

Open Research Online

The Open University's repository of research publications
and other research outputs

Regulation of botulinum toxin complex formation in *Clostridium botulinum* type A NCTC 2916

Thesis

How to cite:

Davis, Tom Owen (1999). Regulation of botulinum toxin complex formation in *Clostridium botulinum* type A NCTC 2916. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1998 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:
<http://dx.doi.org/doi:10.21954/ou.ro.0000e190>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

REGULATION OF BOTULINUM TOXIN COMPLEX

FORMATION IN *Clostridium botulinum*

TYPE A NCTC 2916

bwew
TOM DAVIS

A thesis in partial fulfilment of the requirements of the Open University for the degree
of Doctor of Philosophy.

August 1998

Centre for Applied Microbiology and Research

AUTHOR NO: P9276792

DATE OF SUBMISSION: 28 AUGUST 1998

DATE OF AWARD: 5 JANUARY 1999

DECLARATION

I declare that the research presented in this thesis is my own work, except where otherwise stated, and has not been submitted elsewhere for a research degree.

. Tom Davis

CONTENTS

	Page	
List of contents	i	
List of figures	xiii	
List of tables	xv	
Abbreviations	xvi	
Abstract	xix	
Acknowledgements	xx	
CHAPTER 1	Introduction	1
CHAPTER 2	Materials and Methods	48
CHAPTER 3	Genetic Characterisation of a Silent Type B Neurotoxin Gene From <i>C. botulinum</i> NCTC 2916	70
CHAPTER 4	Molecular Characterisation of the Neurotoxin Operon Found in Association with the Type B Neurotoxin Gene in <i>C. botulinum</i> NCTC 2616	109
CHAPTER 5	Development of a Novel Shuttle Vector Capable of Replication in Clostridia and <i>Escherichia coli</i>	146
CHAPTER 6	Gene Transfer Studies in <i>C. botulinum</i> Type A	183
CHAPTER 7	General Discussion	232
REFERENCES		238
APPENDICES		284

CHAPTER 1 INTRODUCTION	1
1.1 BOTULISM	2
1.2 THE CLOSTRIDIA	2
1.2.1 The genus	2
1.2.2 Solventogenic clostridia	4
1.2.2.1 Clostridial solvent production	4
1.2.3 Pathogenic clostridia	4
1.2.3.1 <i>C. tetani</i>	5
1.2.3.2 <i>C. perfringens</i>	5
1.2.3.3 <i>C. septicum</i>	6
1.2.3.4 <i>C. difficile</i>	7
1.2.3.5 Distribution of clostridial toxins	7
1.3 BOTULISM AND BOTULINAL NEUROTOXINS	8
1.3.1 Intoxication	8
1.3.1.1 Food poisoning	8
1.3.1.2 Wound botulism	9
1.3.1.3 Infant botulism	9
1.3.2 <i>C. botulinum</i> serotypes	10
1.3.3 Physiology of toxin production	11
1.3.4 Toxin complexes	13
1.3.4.1 Evidence for toxin complexes	13
1.3.4.2 Progenitor toxin complex composition	14
1.3.4.3 Antigenic cross-reactivity of toxin complexes	15
1.3.4.3 Neurotoxin	16
1.4 STRUCTURE~FUNCTION ANALYSIS OF CLOSTRIDIAL NEUROTOXINS	16
1.4.1 Functions of the neurotoxin	16
1.4.2 Neurotoxin structure	17
1.4.3 Neuron binding	17
1.4.4 Internalisation	19

1.4.5	Vesicle membrane translocation	19
1.4.6	Intracellular activity	20
1.4.7	Functions of the non-toxic/haemagglutinin components	21
1.5	GENETICS OF NEUROTOXIN PRODUCTION	22
1.5.1	Location of neurotoxin genes	22
1.5.2	Neurotoxin gene sequences	25
1.5.3	Genetic location of components from the neurotoxin complex	26
1.6	DETECTION OF <i>C. botulinum</i>	27
1.6.1	Immunological based detection methods	27
1.6.2	DNA hybridisation detection methods	28
1.7	CLINICAL APPLICATIONS OF NEUROTOXINS	29
1.7.1	Clinical applications of neurotoxins	29
1.8	CLOSTRIDIAL GENETICS	30
1.8.1	Transcriptional factors	30
1.8.2	Transcriptional signals	31
1.8.3	Translational signals	32
1.9	GENETIC MANIPULATION OF CLOSTRIDIA	32
1.9.1	Genetic elements	33
1.9.1.1	Indigenous plasmids	33
1.9.1.2	Heterologous plasmids	34
1.9.1.3	Conjugative plasmids	35
1.9.2	Transposons	35
1.9.2.1	Conjugative transposons	35
1.9.2.2	Non-conjugative Transposons	36
1.9.3	Insertion sequences	37
1.10	GENE TRANSFER TECHNIQUES IN CLOSTRIDIA	37
1.10.1	Natural transformation	37
1.10.2	Transformation by chemical permeabilisation	37

1.10.3	Protoplast transformation	38
1.10.4	Conjugative transfer	38
1.10.5	Electrotransformation	39
1.11	PATHOGENIC BACTERIA	40
1.11.1	Virulence determinants	40
1.11.1.1	Toxins	40
1.11.1.2	Adherence factors	42
1.11.1.3	Invasive factors	43
1.11.1.4	Regulation of virulence factors	44
1.11.1.5	Pathogenicity islands~ inheritance and dissemination of virulence	46
1.12	RESEARCH AIMS	47
CHAPTER 2 MATERIALS AND METHODS		48
2.1	MATERIALS	49
2.1.1	CHEMICALS AND KITS	49
2.1.2	NUCLEIC ACIDS	50
2.1.3	BACTERIAL STRAINS	52
2.2	METHODS	53
2.2.1	STORAGE AND MAINTENANCE OF BACTERIAL STRAINS	53
2.2.2	PURIFICATION OF DNA	53
2.2.2.1	Small scale plasmid isolation from <i>E. coli</i>	53
2.2.2.2	Medium scale plasmid isolation from <i>E. coli</i>	54
2.2.2.3	Large scale plasmid isolation	54
2.2.2.4	Isopycnic centrifugation	54

2.2.2.5	Small scale bacteriophage isolation	55
2.2.2.6	Large scale bacteriophage DNA isolation	55
2.2.2.7	Isolation of DNA from clostridial strains	55
2.2.2.8	Phenol Extraction	56
2.2.2.9	Column Purification	56
2.2.2.10	Ethanol precipitation	57
2.2.2.11	Analysis of DNA	57
2.2.3	TRANSFORMATION OF BACTERIA	57
2.2.3.1	Preparation of competent <i>E. coli</i>	57
2.2.3.2	Transformation of competent <i>E. coli</i> cells with plasmid DNA	58
2.2.3.3	Transfection of competent <i>E. coli</i> with bacteriophage DNA	58
2.2.3.4	Transformation of <i>C. beijerinckii</i> strains	58
2.2.4	DETERMINATION OF PLASMID STABILITY	59
2.2.4.1	Structural stability in <i>C. botulinum/C. beijerinckii</i>	59
2.2.4.2	Segregational stability in <i>C. beijerinckii</i>	59
2.2.5	ELECTROPHORESIS OF NUCLEIC ACIDS	59
2.2.5.1	Horizontal slab gel electrophoresis of DNA	59
2.2.5.2	Horizontal slab gel electrophoresis of RNA	60
2.2.5.3	Recovery of DNA from agarose gels	60
2.2.5.4	Denaturing polyacrylamide gel electrophoresis	60
2.2.6	MODIFICATION OF DNA WITH ENZYMES	61
2.2.6.1	Restriction endonucleases	61
2.2.6.2	Klenow Fragment	61
2.2.6.3	Calf Intestinal Alkaline Phosphatase	61
2.2.6.4	T4 polynucleotide ligase	
2.2.7	IMMOBILISATION OF NUCLEIC ACIDS ON NYLON MEMBRANES	62
2.2.7.1	Southern blotting	62
2.2.7.2	Northern Blotting	62

2.2.7.3	<i>In-situ</i> colony hybridisation	62
2.2.8	PREPARATION OF DNA PROBES	63
2.2.8.1	Radio-labelling of DNA probes	63
2.2.8.2	Radio-labelling of DNA size markers	63
2.2.9	HYBRIDISATION	63
2.2.9.1	Hybridisation of probes to nucleic acids	63
2.2.9.2	Autoradiography	64
2.2.10	PURIFICATION AND ANALYSIS OF RNA	64
2.2.10.1	Culture of strains	64
2.2.10.2	Extraction of total RNA from <i>C. botulinum</i> strains	64
2.2.10.3	Analysis by Primer Extension	65
2.2.10.4	Analysis by Primer Extension-Blotting	66
2.2.11	DNA SEQUENCE ANALYSIS	66
2.2.11.1	Primers	66
2.2.11.2	Plasmid template	66
2.2.11.3	M13 bacteriophage template	66
2.2.11.4	Manual sequencing reactions	67
2.2.11.5	Automated DNA sequencing reactions	67
2.2.11.6	Analysis of data	67
2.2.12	POLYMERASE CHAIN REACTION (PCR)	67
2.2.12.1	Oligonucleotide primers	67
2.2.12.2	Template preparation	68
2.2.12.3	Reagents and conditions	68
2.2.12.4	Cloning PCR products	69

CHAPTER 3	GENETIC CHARACTERISATION OF A SILENT	70
	TYPE B NEUROTOXIN GENE FROM <i>C. botulinum</i>	
	NCTC 2916.	
3.1	INTRODUCTION	71
3.2	STRATEGY	71
3.3	CLONING OF THE NEUROTOXIN LOCUS	71
3.3.1	Cloning of a DNA fragment 5' to the silent neurotoxin gene	71
3.3.2	Investigation of the 5' coding sequence	71
3.3.3	Creation of a genomic restriction map of the silent neurotoxin gene N-terminal coding region.	76
3.3.4	Creation of a genomic restriction map of the silent neurotoxin gene	76
3.3.5	Cloning the silent neurotoxin gene	77
3.3.6	Nucleotide sequence analysis of silent neurotoxin gene in <i>C. botulinum</i> NCTC 2916.	77
3.3.7	Comparison of the silent neurotoxin gene in <i>C. botulinum</i> NCTC 2916 with the published proteolytic and non-proteolytic type B neurotoxin sequences.	78
3.3.7.1	Nucleotide sequence	78
3.3.7.2	Amino acid sequence	85
3.3.8	Comparison of the 5 prime region	87
3.4	TRANSCRIPTIONAL MAPPING OF THE SILENT B GENE USING NORTHERN BLOTTING AND PRIMER TRANSCRIPT ANALYSIS	87
3.4.1	Northern analysis	88
3.4.2	Primer transcript analysis	88
3.4.3	A novel modification for primer transcript analysis	93
3.5	DISCUSSION	93
3.5.1	A silent neurotoxin gene	93
3.5.2	Genetic anomalies of the silent type B neurotoxin gene	97

3.5.3	Interpretation of the Northern blotting	98
3.5.4	Interpretation of primer transcript analysis	98
3.5.5	Translational coupling of the NTNH and BoNT/B transcripts?	106
3.5.6	Function of the silent BoNT/B 'pseudogene'	106
3.5.7	Reversion to an active neurotoxin?	107
3.5.8	Why do type A group I <i>C. botulinum</i> strains commonly contain type B pseudogenes?	108
CHAPTER 4	MOLECULAR CHARACTERISATION OF THE	109
	NEUROTOXIN OPERON FOUND IN ASSOCIATION	
	WITH THE TYPE B NEUROTOXIN GENE IN <i>C. botulinum</i>	
	NCTC 2916.	
4.1	INTRODUCTION	110
4.2	RESULTS	110
4.2.1	Characterisation of the silent type B neurotoxin complex locus	110
4.2.1.1	Evidence of neurotoxin complex genes accompanying the silent type B neurotoxin gene.	110
4.2.1.2	Cloning strategy	111
4.2.1.3	ORF 1	114
4.2.1.4	ORF 2	114
4.2.1.5	ORF 3	115
4.2.1.6	ORF 4	119
4.2.1.7	Origin of the neurotoxin complex sequences 5' to BoNT/A	120
4.2.1.8	ORF 5	122
4.2.1.9	ORF 6	123
4.2.2	Repetitive sequences	124
4.2.3	Transcripts	124
4.2.4	Transcriptional signals	127
4.3	DISCUSSION	131
4.3.1	Haemagglutinin genes	131

4.3.2	ORF 3 ~ A putative transcriptional factor?	134
4.3.2.1	Similarity to the σ^{70} family of proteins?	134
4.3.2.2	The helix-turn-helix motif	135
4.3.2.3	Parallels with the <i>uviA</i> and ORFtxe1 loci	135
4.3.2.4	A member of the FixJ family?	136
4.3.2.5	A member of the ECF σ family?	137
4.3.3	Promoter structure	139
4.3.4	The NTNH proteins	142
4.3.5	Recombination	143
4.3.6	What advantage does a silent neurotoxin gene confer to <i>C. botulinum</i> NCTC 2916?	144
CHAPTER 5	DEVELOPMENT OF A NOVEL SHUTTLE VECTOR CAPABLE OF REPLICATION IN CLOSTRIDIA AND <i>Escherichia coli</i>.	146
5.1	INTRODUCTION	147
5.1.1	Why develop another vector system?	147
5.1.2	Important features in vector design	148
5.2	ANALYSIS OF pCB102	150
5.2.1	Nucleotide sequence analysis	150
5.2.2	Analysis of the pCB102 clone pCB4	157
5.3	DETERMINATION OF THE MINIMAL REPLICON OF pCB102	157
5.3.1	Deletion analysis of pCB4	157
5.3.2	Functional analysis of pCB4 and its deletion derivatives	160
5.3.3	Segregational stability	163
5.4	CONSTRUCTION OF AN <i>E. coli</i> -CLOSTRIDIAL SHUTTLE VECTOR SYSTEM	163
5.4.1	Vector construction	164
5.4.2	Functional analysis of pMTL540E and pMTL541E	165

5.5	CONSTRUCTION OF A CLOSTRIDIAL <i>IN-VIVO</i> EXPRESSION SYSTEM	165
5.6	DISCUSSION	167
5.6.1	Inspection of the minimal replicon	167
5.6.1.1	Analysis of ORFs	167
5.6.1.2	Analysis of secondary structure	168
5.6.1.3	Mechanism of replication encoded by pCB102	169
5.7	FUNCTIONAL DEDUCTIONS FOR pCB102	175
5.7.1	Evidence for a Conjugal operon?	175
	CHAPTER 6 GENE TRANSFER STUDIES IN <i>C. botulinum</i> TYPE A	183
6.1	INTRODUCTION	184
6.1.1	The objective of gene transfer	184
6.2	ELECTROTRANSFORMATION OF <i>C. botulinum</i> NCTC 2916	185
6.2.1	Electrotransformation	185
6.2.2	Validation of putative transformants	186
6.2.3	Isolation of plasmid DNA from putative transformant colonies	186
6.2.4	Large scale plasmid isolation from putative transformant colonies	187
6.2.5	Optimisation of gene transfer in <i>C. botulinum</i> NCTC 2916	188
6.2.5.1	Oxygen contamination	188
6.2.5.2	Estimation of cell viability	188
6.2.5.3	Effect of field strength on transformation efficiency in <i>C. botulinum</i> NCTC 2916	192
6.2.5.4	Effect of DNA concentration on the transformation efficiency of <i>C. botulinum</i> NCTC 2916	194
6.2.5.5	Other parameters	195
6.2.5.6	Electrotransformation of <i>C. botulinum</i> NCTC 2916 with pMTL540E	195

6.3	ELECTROTRANSFORMATION OF <i>C. botulinum</i> ATCC 3502	196
6.3.1	Electrotransformation	196
6.3.2	Analysis of the putative transformants	197
6.3.3	Large scale plasmid isolation from <i>C. botulinum</i> ATCC 3502 transformants	198
6.3.4	Electrotransformation of <i>C. botulinum</i> ATCC 3502 with passaged DNA	198
6.3.5	Effect of a cell wall modification agent on transformation efficiency in <i>C. botulinum</i> ATCC 3502	200
6.4	CHARACTERISATION OF A PLASMID FROM <i>C. botulinum</i> NCTC 2916	202
6.4.1	Justification for cloning a plasmid indigenous to a group I, <i>C. botulinum</i> strain	202
6.4.2	Restriction endonuclease mapping of the putative plasmid from <i>C. botulinum</i> NCTC 2916	202
6.4.3	Cloning a cryptic plasmid from <i>C. botulinum</i> NCTC 2916	203
6.4.4	Determination of the complete nucleotide sequence of the indigenous plasmid from <i>C. botulinum</i> NCTC 2916	203
6.4.4.1	Nucleotide sequence analysis	203
6.4.4.2	Analysis of putative ORFs	204
6.4.4.3	Non-coding sequences	218
6.5	DISCUSSION	222
6.5.1	Transformation of group I <i>C. botulinum</i> type A strains	222
6.5.1.1	Evaluation of gene transfer in <i>C. botulinum</i> type A	222
6.5.1.2	Electroporation parameters	223
6.5.1.3	Host encoded factors	223
6.5.1.4	Incompatibility	226
6.5.1.5	Conclusion	226
6.5.2	Theoretical characterisation of pBP1	226
6.5.2.1	Method of replication	226
6.5.2.2	Pre protein function	228
6.5.2.3	Cassette structure	229
6.5.2.4	Natural distribution of pBP1	230

6.5.2.5	Potential uses	231
CHAPTER 7 GENERAL DISCUSSION		232
7.1	INTRODUCTION	233
7.2	THE TYPE B NEUROTOXIN GENE IN <i>C. botulinum</i> NCTC 2916 TYPE A	233
7.3	RECOMBINATION IN <i>C. botulinum</i>	235
7.4	REGULATION OF NEUROTOXIN COMPLEX REGULATION IN <i>C. botulinum</i>	235
7.5	DEVELOPMENT OF A GENETIC SYSTEM~CRITICAL EVALUATION	236
REFERENCES		238
APPENDICES		284
APPENDIX I	Buffers and stock solutions	284
APPENDIX II	Media	290
APPENDIX III	Annotated nucleotide sequence of neurotoxin complex operons 5' to the silent type B neurotoxin in <i>C. botulinum</i> NCTC 2916	293
APPENDIX IV	Comparative alignment of the NTNHI gene from the type B neurotoxin locus in <i>C. botulinum</i> type A, strain NCTC 2916, with NTNHI genes from other serotypes	302
APPENDIX V	Oligonucleotide primers	319

FIGURES

Figure number		Page
3.1	Selective PCR amplification from the silent type B neurotoxin locus	73
3.2	Schematic representation of the strategy used to clone genomic fragments of the type B neurotoxin locus	75
3.3	Comparitive alignment of the silent BoNT/B gene	79
3.4	Comparitive alignment of the type B neurotoxin amino acid sequences	83
3.5	Comparitive alignment of a clostridial neurotoxin motif absent in BoNT/B (silent)	86
3.6a	Northern blot of <i>C. botulinum</i> total RNA probed with a type A neurotoxin specific probe	89
3.6b	Northern blot of <i>C. botulinum</i> total RNA probed with a type B neurotoxin specific probe	90
3.7	Detection of 5' mRNA transcripts by primer extension	91
3.8	Development of a primer extension blotting technique	94
3.9	Detection of 5' mRNA ends by primer extension-blotting	95
3.10	Nucleotide sequence discrepancy in the silent BoNT/B neurotoxin sequence	99
3.11	Alignment of the intergenic regions between botulinal neurotoxin serotypes	101
3.12	Hypothetical mechanism of NTNH-Neurotoxin transcript degradation	103
3.13	Hypothetical function of the stalled ribosome	105
4.1	Cloning strategy for the type A and type B neurotoxin loci from <i>C. botulinum</i> NCTC 2916	112
4.2	Prediction of H-T-H motifs	117
4.3	Alignment of ORF 3 homologues with SigX	118
4.4	Comparitive alignments of the HA34 haemagglutinin component of <i>C. botulinum</i> NCTC 2916.	121
4.5	Alignment of HA70 derivatives with the <i>C. perfringens</i> CPE enterotoxin	125
4.6	Translational signals from neurotoxin locus genes from <i>C. botulinum</i> NCTC 2916	126

4.7a	Generating a genomic restriction map of the type B neurotoxin locus in <i>C. botulinum</i> NCTC 2916	128
4.7b	Generating a genomic restriction map of the type B neurotoxin locus in <i>C. botulinum</i> NCTC 2916	129
4.8	Clostridial neurotoxin promoters and other cognate promoters	130
5.1	Schematic map of plasmid pCB102	151
5.2	Homologues of ORFA encoded on the <i>Clostridium butyricum</i> plasmid pCB102	152
5.3	Alignment of ORF I from pCB102 with site-specific single strand endonucleases.	153
5.4	Alignment of the putative origin of transfer from pCB102	156
5.5	Comparative alignment of hydrophilicity profiles for ORF D of pCB102 and putative homologues	158
5.6	Alignment of ORF D (pCB102) with putative homologues	159
5.7	Deletion analysis of pCB4	161
5.8	Frame-shift mutation of ORF H	162
5.9	Putative regulatory structures within the minimal replicon of pCB102	170
5.10	Secondary structure of pCB102 replicon RNA species	171
5.11	Construction of <i>E. coli</i> -clostridial shuttle vectors	179
5.12	Construction of an expression vector system	180
5.13	Secondary structure and hydrophobicity profiles for putative ORF K of pCB102	181
6.1	Restriction endonuclease analysis plasmid DNA isolated from <i>C. botulinum</i> NCTC 2916	189
6.2	Restriction endonuclease mapping of plasmid DNA isolated from <i>C. botulinum</i> ATCC 3502	199
6.3	Complete annotated nucleotide sequence of pBP1	205
6.4	Schematic map of pBP1	211
6.5	Comparative alignment of the pBP1 ORF 1 with pC194 type replication proteins	212
6.6	Putative signal peptides of ORF 2 and ORF 3 from pBP1	216
6.7	Comparative alignment of ORF 8 from pBP1 with Pre/Mob proteins	219

TABLES

Table Number		Page
3.1	Distribution of nucleotide substitutions within the codons of three pairs of neurotoxin genes	85
3.2	Examples of mRNA stability influenced by Shine-Dalgarno type sequences	104
4.1	Weight matrix for the prediction of helix-turn-helix motifs	117
4.2	Calculation of helix-turn-helix probability	117
5.1	Estimation of plasmid segregational stability	164
6.1	Correlation of cell viability and transformation efficiency	190
6.2	Effect of electric field strength on transformation efficiency in <i>C. botulinum</i> NCTC 2916 experiment 1	192
6.3	Effect of electric field strength on transformation efficiency in <i>C. botulinum</i> NCTC 2916 experiment 2	193
6.4	Effect of DNA concentration on transformation efficiency in <i>C. botulinum</i> NCTC 2916	194
6.5	Electroporation of <i>C. botulinum</i> ATCC 3502 with pMTL540E	197
6.6	Electroporation of <i>C. botulinum</i> ATCC 3502 with plasmid DNA from different sources	200
6.7	Effect of a cell wall modification agent on gene transfer frequency in <i>C. botulinum</i> ATCC 3502	201
6.8	Putative ORFs of pBP1	214

ABBREVIATIONS

%	Percent
σ	Sigma factor
Ω	Resistance
"	Inch
°C	Degrees Centigrade
A	Adenine
ADP	Adenosine Diphosphate
Amp ^R	Ampicillin resistance
AMV	Avian Myeloblastosis Virus
ATCC	American Type Culture Collection
BoNT	Botulinum Neurotoxin
bp	Base pair
C	Cytosine
CAMR	Centre for Applied Microbiology and Research
cDNA	Complementary Deoxyribonucleic Acid
CFU	Colony Forming Unit
CIP	Calf Intestinal Alkaline Phosphatase
cm	Centimetre
CMC	Cooked Meat Carbohydrate
CPE	<i>Clostridium perfringens</i> Enterotoxin
dATP	Deoxyadenosine 5' Triphosphate
dCTP	DeoxyCytosine 5' Triphosphate
DEAE	Diethylaminoethyl
dGTP	Deoxyguanosine 5' Triphosphate
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
dNTPs	Deoxyribonucleoside 5' Triphosphates
ds	Double stranded
dso	Double stranded origin
dTTP	Deoxythymine 5' Triphosphate
ELISA	Enzyme-Linked Immunosorbent Assay
Erm ^R	Erythromycin resistance
F	Farad
G	Guanine
g	Gram
G _{T1B}	N-acetylneuraminylgalactosyl-N-acetylgalactosaminy -[N-acetylneuraminy-N-acetylneuraminy]-galactosyl glucosylceramide
h	Hour
H-T-H	Helix-Turn-Helix
HA	Haemagglutinin
IPTG	Isopropyl-thio- β -galactoside
kb	Kilobase
kDa	Kilodalton
V	Volt
l	Litre
M	Molar
mg	Milligram

min	Minute
ml	Millilitre
MLS ^R	Macrolide-Lincosamide-Streptogramin Resistance
mM	Millimolar
mRNA	Messenger Ribonucleic Acid
NCIMB	National Collection of Industrial and Marine Bacteria
NCTC	National Collection of Type Cultures
ng	Nanogram
nm	Nanometre
NTNH	Nontoxic-NonHaemagglutinin
O.D. _n	Optical density at wavelength n
ORF	Open Reading Frame
p	Pico
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PFGE	Pulsed Field Electrophoresis
pI	Isoelectric point
r.p.m	Revolutions per minute
rATP	Adenosine Triphosphate
RBC	Red Blood Cell
RC	Rolling Circle
RF	Replicative Form
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
S	Svedberg (sedimentation coefficient)
SD	Shine Dalgarno
SDS	Sodium Dodecyl Sulphate
sec	Second
SNAP-25	Synaptosomal associated protein of 25 kDa
ss	Single stranded
sso	Double stranded
T	Thymine
TeNT	Tetanus Neurotoxin
U	Uracil
UV	Ultraviolet
V	Volt
v/v	Volume to volume ratio
VAMP	Vesicle associated membrane protein
vol.	Volume
w/v	Weight to volume ratio
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
oc	Open circular
ccc	Covalently closed circular

Amino Acids

Amino Acid	Single letter code
Alanine	A
Arginine	R
Asparagine	N
Aspartic Acid	D
Cysteine	C
Glutamic Acid	E
Glutamine	Q
Glycine	G
Histidine	H
Isoleucine	I
Leucine	L
Lysine	K
Methionine	M
Phenylalanine	F
Proline	P
Serine	S
Threonine	T
Tryptophan	W
Tyrosine	Y
Valine	V

ABSTRACT

REGULATION OF BOTULINUM TOXIN COMPLEX FORMATION IN

Clostridium botulinum TYPE A NCTC 2916

by Tom O. Davis

Genomic DNA fragments encoding the silent type B neurotoxin gene from *Clostridium botulinum* NCTC 2916 have been cloned and the complete nucleotide sequence determined. The translated sequence revealed that the gene encoded a neurotoxin which was closely related to type B neurotoxin genes from Group I *Clostridium botulinum*. However among the nucleotide sequence differences, a G to T transition has interrupted the coding sequence with the formation of a stop codon. In addition the deletion of an adenine residue has resulted in a frame-shift mutation.

Analysis of the DNA sequence contiguous with the silent type B neurotoxin gene revealed the presence of a gene encoding a Nontoxic-Nonhaemagglutinin protein which appears to share a bicistronic mRNA transcript with the type B neurotoxin gene. In the reverse orientation, the partial sequence of a gene encoding a haemagglutinin protein was found, typical of type A and B botulinal neurotoxin complexes. Separating the genes encoding the components of the neurotoxin complex was a gene of 178 amino acids which possessed features commonly associated with transcriptional factors.

To facilitate the *in vivo* study of botulinal neurotoxin complex regulation, a gene transfer system using clostridial components has been developed. The minimal replicon of the cryptic plasmid pCB102 from *Clostridium butyricum* NCIB 7423 was located to 1.6 kb DNA fragment by deletion analysis, enabling the identification of hitherto undiscovered putative ORFs and secondary structures, consistent with a replicative function. The replicon has been incorporated in to a number of *Escherichia coli* vectors resulting in a versatile series of shuttle vectors which have demonstrated high structural and segregational stabilities in a heterologous host *Clostridium beijerinckii* NCIMB 8052. Gene transfer of a Group I *Clostridium botulinum* type A strain was demonstrated with a representative pCB102-derived shuttle vector, pMTL540E.

In addition, a 5.9 kb plasmid indigenous to *C. botulinum* NCTC 2916 was cloned and the complete nucleotide sequence determined. Eight putative ORFs have been identified, including a putative replication protein and recombinase.

ACKNOWLEDGEMENTS

I would like to thank Dr Ian Henderson and Dr Nigel Minton for their excellent supervision and intellectual contribution during the course of this study, and for finding the time to proof read this thesis. I would also like to acknowledge the valuable advice and discussion with my colleagues in CAMR, both past and present.

CHAPTER 1

INTRODUCTION

1.1 Botulism

Bacteria which prey upon the human being take many forms, but among these, *Clostridium botulinum* holds a particularly fearful reputation due to the extreme potency of its neurotoxins and invariably fatal consequences. There are now seven serological groupings of the neurotoxin (A-G) and within this framework a number of distinct subgroups exist. Irrespective of the serotype, botulinal neurotoxins occur as part of a heterogeneous protein complex which commonly enter the body by ingestion, where upon they eventually effect their toxic activity in the synapses of neuromuscular junctions, causing flaccid paralysis and death. No single species is responsible for the production of these protein toxins which have been attributed to at least six lineages of the genus *Clostridium*. Whilst there is a limited understanding of the composition of the neurotoxin complex, information on the mechanisms by which the various components are regulated is virtually nonexistent. To further our understanding of neurotoxin complex regulation, the components must first be characterised at the genetic level. An essential feature of this approach is the development of a genetic system by which studies can be conducted *in vivo*.

1.2 The Clostridia

1.2.1 The genus

In the latter part of the nineteenth century, the discovery of a number of anaerobic fermentative spore forming bacteria prompted the concept of the genus *Clostridium* by Prazmowski (1880). However it was Trecul, who in 1863 devised the term *Clostridium* which is derived from kloster, the greek word for spindle and the diminutive suffix 'ium', thus describing a 'small spindle'. It was at this time that organisms resembling many of the medically and biotechnologically important clostridia were first described including

Clostridium butyricum (1880), *Clostridium tetani* (1884), *Clostridium septicum* (1887), *Clostridium perfringens* (1889), *Clostridium difficile* (1893) and *Clostridium pasteurianum* (1895), indeed *C. botulinum* was described by van Ermengen in 1896 (see Morris, 1993).

Clostridia are distinguished by their ability to form endospores, are generally Gram-positive rods and on the whole are obligate anaerobes. They are believed to metabolise a wide range of substances including carbohydrates, alcohols, amino acids, purines, steroids and other organic compounds (Cato *et al.*, 1986). Some species fix atmospheric nitrogen but none are capable of dissimilatory sulfate reduction. The wide repertoire of metabolic capabilities of this genus underlies a very diverse spectrum of organisms which are collectively grouped on a rather broad basis of shared phenotypic traits. The classical taxonomic methods of distinguishing these organisms (i.e., Cato *et al.*, 1986) have recently been superseded by DNA based studies in which the nucleotide sequence divergence of 16S rRNA genes is correlated with phylogenetic relationships. On this basis, previous groupings have been refined to produce 19 separate phylogenetic clusters (Collins *et al.*, 1994). Cluster 1 which accommodates at least half of the currently known clostridia includes the *C. butyricum* type strain, the neurotoxigenic clostridia and *C. perfringens*, but not *C. difficile*. Members of this group retain the original 'clostridial' status. The association of clostridial species with a variety of taxa has generated phylogenetic clusters containing combinations of organisms with and without the original 'clostridial criteria,' e.g., spore-formers and non-spore formers, Gram-negative and Gram-positive, aerobes and obligate anaerobes and DNA over a range of mol % G + C (Collins *et al.*, 1994).

Three broad groups of clostridia have received disproportionate scientific attention due to their etiological association with a number of diseases or biotechnological potential.

These include pathogens and solventogens, and cellulolytic thermophiles which are outside the scope of this review.

1.2.2 Solventogenic clostridia

1.2.2.1 Clostridial solvent production

Some species of clostridia are able to metabolise a wide range of organic compounds into acids, CO₂, H₂ and a range of neutral solvents. It was the ability to produce solvents which gave particular species of clostridia a niche in the interests of twentieth century man. Fermentation of waste biomass from agriculture and industry can be transformed into chemical feed stocks and fuels. One such process developed on an economic scale termed acetone/butanol/ethanol fermentation (ABE fermentation) was economically viable until the 1960's (Jones & Woods, 1989). However to compete with contemporary chemical processes, microbial solvent production must become more efficient, specifically by alterations in the profile and quantities of solvents produced by the clostridia. To achieve this goal, the genetic basis of solventogenesis has been studied in detail resulting in the cloning and analysis of genes encoding many key metabolic enzymes (reviewed by Girbal & Soucaille, 1998). The parallel development of genetic systems to study and manipulate these genes has been an essential element in this field, which also provides useful genetic tools for work with other clostridia (reviewed by Minton, Young and Durre in 'The Clostridia and Biotechnology' 1993).

1.2.3 Pathogenic clostridia

Amongst the clostridia, many species are pathogens of both animals and man, causing a range of conditions including specific neurological disorders, enteric infections and

necrotising tissue damage. Due to the large content of this subject, several principal clostridial pathogens will be considered with which *C. botulinum* can be compared.

1.2.3.1 *C. tetani*

Disease caused by *C. tetani* is more widespread than that caused by *C. botulinum*, being endemic in 90 developing countries (Arnon, 1997). Infection initiates by the introduction of spores into areas of tissue trauma in which a low redox potential is generated, typically occurring in deep wounds, umbilical cords of neonates, and post-partum, post-abortion or post-surgical infections (Arnon, 1997). *C. tetani* is known to produce a hemolysin. However, its principal toxin is a neurotoxin which is related to the botulinal neurotoxins. The neurotoxin is released by autolysis at the site of infection from where it migrates in retrograde motion through the motor neurons to inhibitory interneurons of the brain stem and spinal cord. Here, tetanus neurotoxin acts as an endoprotease, cleaving the synaptic vesicle docking protein synaptobrevin, a substrate shared with some other clostridial neurotoxins (Montecucco & Schiavo, 1993). In stark contrast to the flaccid paralysis caused by botulism, tetanus is an acute spastic paralysis. This is because their respective neurotoxins act at distinct sites within the nervous system and hence the profoundly different pathologies are presented.

1.2.3.2 *C. perfringens*

In contrast to *C. botulinum* and *C. tetani*, which generally elaborate a single neurotoxin, most other clostridial pathogens elicit disease with a multifaceted assembly of toxic molecules. *C. perfringens* is such an organism.

Different isolates of *C. perfringens* are essentially identical. Where strain variation does occur it is due to the presence or absence of a limited number of genes, particularly

virulence determinants which in some cases can be either genomically or extrachromosomally located (Cole & Canard, 1997). *C. perfringens* is implicated in a number of human and animal diseases which necessitate different virulence factors, for instance a form of food poisoning is mediated by a single enterotoxin (CPE), whilst gas gangrene requires two major cytolytic factors, phospholipase C which is a phospholipase-sphingomyelinase (PLC) and perfringolysin O, a pore-forming cytotoxin. Other virulence factors are thought to include a collagenase, proteases, a hyaluronidase (*nagH*), neuramidase (*nanH*), and a so called “epsilon toxin” of unknown activity. Necrotic enteritis associated with type B and C *C. perfringens* strains is thought to be mediated by a ‘membrane active’ β -toxin and often occurs in malnourished people, particularly in combination with diets which are low in protein and high in trypsin inhibitors which may protect the protein toxin (Stevens, 1997). The large number of toxins produced by *C. perfringens* reflects the method by which this organism establishes its niche and subsequently exploits the potential of the host tissues.

1.2.3.3 *C. septicum*

C. septicum is associated with non-traumatic or spontaneous gas gangrene. Disease is frequently enabled due to other predisposing factors, allowing establishment and subsequent access to the host blood stream. Tissues are colonised through the action of at least four toxins including a lethal hemolytic necrotising α -toxin, a DNase (β -toxin), a hyaluronidase (γ -toxin) and hemolysin (δ -toxin) (Stevens, 1997). In common with *C. perfringens*, *C. septicum* possess a number of virulence factors which facilitate a pathogenic life style.

1.2.3.4 *C. difficile*

C. difficile possess two cytotoxic proteins designated A and B. Although toxin B has the more potent cytotoxic activity, toxin A also has enterotoxic activity and is responsible for the initial intestinal damage in pseudomembranous colitis (Moncrief *et al.*, 1997). Toxin A and B are the largest known bacterial toxins and share 63.1 % similarity. Both toxins are thought to be endocytotically internalised (Moncrief *et al.*, 1997). Although enzymic in nature, *C. difficile* cytotoxins do not degrade host tissues but confer a subtle modification to their substrates, the Rho family, proteins which participate in regulating actin polymerization and, therefore, the integrity of the eukaryotic cytoskeleton (Aktories *et al.*, 1997).

Toxin A and B are not unique, since both *Clostridium sordellii* and *Clostridium novyi* possess large toxins which appear to be genetically related to the *C. difficile* toxins, indeed the two toxins found in *C. sordellii* roughly equate to toxin A and B of *C. difficile*. Recent evidence suggest that although these proteins may have a common ancestry, the range of members of the Ras superfamily which are modified is distinct for each toxin (Aktories *et al.*, 1997) and the pathological effect to the host cell are, therefore, different.

1.2.3.5 Distribution of clostridial toxins

A number of toxin types occur throughout the pathogenic clostridia, these include thiol-activated cytolysins which occur in most pathogenic clostridia (Tweten, 1997) including *C. perfringens* (perfringolysin O), *C. tetani* (tetanolysin) and *C. botulinum* (botulinolysin), the phospholipases C which occur in *C. perfringens* strains A-E, *Clostridium bifermentans*, *C. novyi*, *C. sordellii*, *Clostridium paraperfringens* and

Clostridium absonum (Titball, 1997), the binary toxins of *C. perfringens* type E and *Clostridium spiroforme*, collagenases of *C. perfringens* (all types) and *Clostridium histolyticum* (Okabe & Cole, 1997) and the sialidases produced by *C. perfringens*, *C. sordellii*, *C. septicum*, *Clostridium chauvoei*, *Clostridium tertium* and particular strains of *C. butyricum* and *Clostridium nexile* (Roggentin & Schauer, 1997). In essence many of these clostridial species seem to have a similar strategy, which consists of potent 'lethal toxins' which are required to inflict debilitating damage to the host (for example, *C. perfringens* type A α -toxin) and a battery of synergistic 'toxins' that specialise in degrading the host tissues and spreading infection, which in the case of *C. perfringens* could include collagenase, hyaluronidase, sialidase and proteases.

1.3 Botulism and botulinal neurotoxins

1.3.1 Intoxication

Botulism appears to affect vertebrate organisms through the action of the neurotoxins (serotypes A-G) and associated non-toxic proteins. Different mammals are intoxicated with varying efficiencies depending on the neurotoxin serotype (Sakaguchi *et al.*, 1988) and the extent to which the neurotoxin is complexed with other proteins (Sugiyama, 1980). The neurotoxic nature of these molecules has also been demonstrated in non-vertebrate organisms that contain neuronal networks (Poulain *et al.*, 1990) indicating that the targets of the neurotoxins must be both ancient and highly conserved.

1.3.1.1 Food poisoning

Ingestion of food contaminated with growing *C. botulinum* is the principal method by which humans are intoxicated. The major source of contamination stems from the incorrect preparation and preservation of home prepared foods (Hatheway, 1995).

Serotypes of *C. botulinum* which are most commonly associated with human poisoning include type A, type B and type E. The prevalence of a particular serotype can be attributed to the type of food, for example, strains of group I serotype A have spores which are resistant to thermal processing used in home canning whilst less thermally resilient strains from group II are frequently encountered in cured meat products (Hatheway, 1995). A condition known as forage poisoning occurs in herbivores such as horses and cattle that ingest fodder contaminated with decomposing organisms which contain *C. botulinum* types C and D, or alternatively from the growth of *C. botulinum* in the feed. Ingestion of toxic material would appear to be a major vector for botulism in the natural environment. The persistence of spores in animals frequently results in contamination of cadavers with *C. botulinum*. Maggots which feed on the rotting flesh ingest neurotoxin which does not harm the maggot and is not degraded. Wild fowl are then particularly prone to botulism by eating maggots from the cadavers, ingesting the toxin and dying, and hence potentiating the cycle. In this manner vast numbers of birds can die (Kalmbach, 1968).

1.3.1.2 Wound botulism

This form of the disease resembles infection by *C. tetani*. Damaged tissue which is deep enough to generate an anaerobic environment enables clostridial spores to germinate (Weber *et al.*, 1993). In contrast to tetanus, wound botulism is rare and may reflect the difficulty with which this group of organisms can colonise the wound environment.

1.3.1.3 Infant botulism

The intestine provides an ideal medium in which anaerobic bacteria can thrive. Adults are not readily colonised by *C. botulinum*. In contrast, infants between 3 and 26 weeks of

age are susceptible to intestinal toxico-infection caused by growth predominantly in the colon (Sugiyama, 1980) following the ingestion of spores. Infant botulism is principally caused by type A and B botulinal organisms and in rare cases toxigenic *C. butylicum* (Aureli *et al.*, 1986) and *C. baratii* (Hall *et al.*, 1985). Evidence suggests that the natural microflora of adults provides a barrier to *C. botulinum* colonisation, however, in instances when this is not present (such as axenic mouse models), individuals such as young infants become susceptible (Moberg & Sugiyama, 1979).

1.3.2 *C. botulinum* serotypes

In contrast to organisms which elaborate the serologically unique tetanus neurotoxin (tetanospasmin), botulinal neurotoxins are produced by a range of clostridia. Neurotoxic molecules were initially distinguished by antibody neutralisation. In 1919 Burke isolated type A and B toxins from cases of human botulism, types C-G were then added in chronological order beginning with type C (1922) and type D (1928) which were isolated from cattle. Later, in 1937, type E was isolated from human botulism with a fish origin and, in 1958, type F was associated with poisoning from a liver paste (Hatheway, 1995). Since this time, an additional serotype (G) was described by Gimenez, who isolated the organism from soil (Gimenez & Ciccarelli, 1970).

Comparative analysis of the respective neurotoxigenic organisms using physical characteristics, such as growth optima, spore heat resistance and phenotypic enzyme analysis, has divided the bacteria into four separate lineages (group I-IV in Hatheway, 1995). Group I are proteolytic and are able to digest casein and other complex proteins, a property not shared by group II organisms, which are nonproteolytic. Organisms which resemble group I may encode type A, B or F neurotoxins and in rare cases, organisms are isolated which produce more than one toxin e.g., combinations are encountered

including A and B in various ratios or F in combination with A or B (Poumeyrol *et al.*, 1983; Hatheway *et al.*, 1981; Hatheway & McCroskey, 1989; Gemenez & Ciccarelli, 1970). Group II organisms can elaborate type E neurotoxin but not type A. Isolates also produce type B and type F but never together.

C. botulinum producing type C or type D neurotoxin have a relatively high optimum growth temperature (37°C - 45°C) and produce a novel combination of volatile organic acids. Organisms which produce type G neurotoxin are distinct from groups I-III as they are neither lipase proficient or saccharolytic (Hatheway, 1995).

The application of DNA technology to taxonomic classification has confirmed the heterogeneity observed between *C. botulinum* species. Sequence analysis of 16S rRNA genes essentially supported the separation of botulinogenic clostridia into four groups and also supported the previous notion that the individual groups were complemented by nontoxigenic counterparts (Collins *et al.*, 1994). However, all neurotoxigenic clostridia, including *C. tetani* and neurotoxigenic forms of *C. butyricum* and *C. baratii* (producing type E and F neurotoxins, respectively (Aureli *et al.*, 1986; Hall *et al.*, 1985), are found within a single cluster of the diverse genus *Clostridium* (Collins *et al.*, 1994).

1.3.3 Physiology of toxin production

A thorough investigation of toxin production by all the representative *C. botulinum* groups does not appear to have been conducted. This may in part be due to the difficulties associated with determining the amounts of toxin present via the mouse bioassay and until recently the absence of defined media (Whitmer & Johnson, 1988). The relationship between the presence of toxin in culture supernatants and autolysis has been documented for some time (Bonventre & Kempe, 1960) and an autolytic enzyme has been characterised in some strains (Takumi *et al.*, 1971; Kawata & Takumi, 1971). It

was shown that intracellular toxin was present throughout the growth phases of batch cultures and its release into the growth media correlated with the extent of autolysis, implicating release by this mechanism. It was noted, however, that disproportionate toxin synthesis occurred “after the period of cell multiplication” (Kindler *et al.*, 1956; Bonventre & Kempe, 1960).

Batch fermentation studies of type A and B strains in complex media were used to assess the contribution of various carbohydrate sources to neurotoxicity in these strains (Bonventre & Kempe, 1959a). Glucose or maltose were found to induce highly neurotoxigenic cultures. In comparison, the presence of glycerol, pyruvate and ribose produced poorly toxigenic cultures. Other carbon sources, including galactose, lactose, xylose and inositol, did not enhance neurotoxicity. However, these results could equally be interpreted as a measure of the activation (proteolytic processing) of pre-existing neurotoxin which could account for the increased neurotoxicity. Further experiments were attempted by the same authors to define the effect of pH and temperature on neurotoxin production (Bonventre & Kempe, 1959b). Toxin was synthesised between pH 5.5-8.3 at a relatively unvarying rate, but at alkaline pH the toxin was unstable. In contrast, temperature was found to alter the toxicity of cultures. Maximum toxicity was detected in cultures between 28 °C and 40 °C, which correlated with both growth and autolysis. Siegel & Metzger (1979; 1980) conducted similar experiments on both type A and type B strains and obtained essentially similar results with each strain. Optimum toxin production was detected at 35 °C with glucose concentrations between 0.5 %-1.5 %. Glucose concentrations above and below these levels were detrimental to toxin production.

Later studies conducted with defined media using proteolytic group I strains of *C. botulinum* demonstrated that both toxin production and protease activity was ‘strongly

influenced by carbon and nitrogen nutrition' (Patterson-Curtis & Johnston, 1989). Arginine is thought to be a precursor of both energy producing and biosynthetic pathways in group I *C. botulinum* (Whitmer & Johnson., 1988). Indeed, the availability of arginine and other nitrogenous compounds appeared to inversely regulate the protease and neurotoxin production in these strains (Patterson-Curtis & Johnston, 1989). In contrast, the presence of casein counteracted this phenomenon. A similar phenomenon involving tryptophan was observed in a type E representative of nonproteolytic group II *C. botulinum* where the amino acid is thought to provide a nitrogen source (Leyer & Johnson, 1990).

In conclusion, in the group I and II *C. botulinum* strains examined, the production of neurotoxins did not appear to directly correlate with growth but may be under regulatory control which is influenced by nutrient availability.

Studies involving *C. tetani* concluded that this neurotoxin was also regulated by nutritional and environmental factors, including histidine rich peptides, nitrogen (glutamate) and the concentration of iron (Johnson, 1997). In summary, relatively little is known about the regulation of clostridial neurotoxin production, particularly at the genetic level. This is in part due to the absence of suitable genetic techniques.

1.3.4 Toxin complexes

1.3.4.1 Evidence for toxin complexes

Initial endeavors to isolate neurotoxin from *C. botulinum* type A resulted in the crystallisation of particles of approximately 900 kDa. However work by Lamanna demonstrated that the preparation was heterogeneous as a haemagglutinin constituent of the 900 kDa moiety could be removed by absorption onto erythrocytes (cited by Sugiyama, 1980). Subsequently a number of chromatographical approaches were then

able to resolve the toxic and haemagglutinin entities (DasGupta & Boroff, 1968) and provide further evidence that botulinal toxins were heterogeneous complexes. The neurotoxin complexes produced by *C. botulinum* strains range in size and are serotype specific. The toxin complexes, or 'progenitor toxins', are classified into four groups. The smallest, at 7 S (150 kDa), contained the toxic moiety and was designated 'S' (for small). However it was not commonly observed in the extracts from any serotype. All serotypes, except G, did produce a progenitor toxin of 12 S (300 kDa), which was designated 'M' (for medium). M toxin could be reversibly dissociated in alkaline conditions to release S toxin and a similar sized non-toxic component. With the exception of serotypes E and F, all other serotypes produced larger forms of progenitor toxin. Serotypes A, B, C and D produced a 16 S (500 kDa) complex, termed L (for large), whilst only serotype A formed an even larger progenitor toxin of 19 S (900 kDa), designated 'LL' (Sakaguchi *et al* 1988). Serotype G progenitor toxins were only isolated as L progenitor complexes (Nukina *et al.*, 1991).

1.3.4.2 Progenitor toxin complex composition

Haemagglutination was described for the type A toxin complex sizes M, L and LL but not S (DasGupta *et al*, 1966; DasGupta *et al.*, 1968). A similar situation was described for other serotypes. In summary, progenitor toxins of 300 kDa (M) were all eventually found to be nonhaemagglutinating, whilst larger complexes (L and LL) were haemagglutinating (Sakaguchi *et al*, 1988). Thus, with the exception of E and F, all serotypes possessed haemagglutinin activity. Sackaguchi and co-workers were then able to demonstrate that increasing sizes of progenitor toxin complexes correlated with enhanced oral toxicity (Sakaguchi *et al.*, 1988). Indeed, in comparison to intraperitoneal administration, it was considered that approximately 10 million times more S toxin was

required to cause toxicity by the oral route. In contrast, a mere 1,500-12,000 fold increase was required for L toxin.

The larger progenitor complexes afford protection to the S toxin in environments which resemble those encountered in the intestine. For example, M or L complexes remain largely recalcitrant to the action of acidic environments at pH 3-4 and by the action of gastric proteases and juices (Sugiyama, 1980). In addition large toxin complexes facilitate the entry of S toxin into the lymph (Sugii *et al.*, 1977).

1.3.4.3 Antigenic cross-reactivity of toxin complexes

The similarity observed between botulinogenic neurotoxins in their clinical symptoms and molecular sizes is to some extent reflected in a number of common antigenic motifs. Monoclonal antibodies raised against the light chain of type E neurotoxin are also able to cross react with type B, C, D and tetanus neurotoxin (TeNT) (Tsuzuki *et al.*, 1988). A reciprocal experiment using synthetic peptides of TeNT provided a similar result (Halpern *et al.*, 1989). More recently, monoclonals have been found which recognise all botulinal neurotoxin serotypes (Dertzbaugh & West, 1996), indicating some sequence homology between all toxins.

The nontoxic components of specific progenitor toxins also appear to be similar. Lamanna noted immunological cross reactivity between the nontoxic component of type A and type B progenitor toxins which were later shown to be of a similar size (cited in Somers & DasGupta, 1991). Likewise, haemagglutinin components of type C and type D progenitor toxins were also immunologically related (Miyazaki *et al.*, 1977). Thus, it appeared that clostridial neurotoxins were related. Furthermore, the non-toxic proteins associated with serologically distinct botulinogenic neurotoxins were also related. Similar methodology has demonstrated that neurotoxins of the same serotype are not identical,

for example, type B neurotoxins from proteolytic and nonproteolytic strains (Shimizu & Kondo, 1973), type F strains (Gimenez & Ciccarelli, 1972) and type C strains (Ochanda *et al.*, 1984) are immunologically variable.

1.3.4.4 Neurotoxin

DEAE-Sephadex chromatography of progenitor toxins of serotype A-F at slightly alkaline pH dissociated the complexes into constituent proteins, among these was the 7 S neurotoxin (Sakaguchi *et al.*, 1988). However, by whatever method used, a 150 kDa protein was attributed with neurotoxic activity in all known clostridial neurotoxin complexes. Analysis of this protein from proteolytic strains A, B and F under reduced denaturing conditions demonstrated the presence of two bands of approximately 50 kDa and 100 kDa (DasGupta & Sugiyama, 1972). However, polypeptides of equivalent size were only obtained from non-proteolytic strains (which do not possess endogenous proteases) following trypsination.

1.4 Structure ~ function analysis of clostridial neurotoxins

1.4.1 Functions of the neurotoxin

In an effort to determine the mechanism by which botulinal neurotoxins poison vertebrates, a plethora of structural and pharmacological investigations have been performed. During the post war period the neurotoxins were shown to block synaptic transmission of cholinergic neurons by preventing acetylcholine release from nerve terminals (Harris & Miledi, 1971). To equate this mode of action with the neurotoxins structure, experiments were conducted to detect cell receptor binding and cell membrane interaction potential as well as an intracellular site of action.

1.4.2 Neurotoxin structure

The analysis of type A neurotoxin revealed that the 150 kDa protein was initially synthesized as a single polypeptide chain. Although it subsequently cleaved into a 50 kDa N-terminal light chain fragment and 100 kDa C-terminal heavy chain fragment, the two chains remained linked by a disulfide bridge (Krieglstein *et al.*, 1994) which was essential for toxicity in vertebrates (Maisey *et al.*, 1988). Among the remaining 7 cysteine residues, a further pair were involved in intra-chain disulphide bridge formation near the C-terminus of the heavy chain (Krieglstein *et al.*, 1994). Peptide sequence analysis of the light chain-heavy chain junction has revealed that in the type A serotype, 10 amino acids which are thought to correspond to a surface exposed loop are excised by endogenous proteases (Krieglstein *et al.*, 1994). The similarities observed between the gross features of the different clostridial neurotoxins are surprisingly well reflected in the high degree of conserved secondary structure motifs predicted along the length of the proteins (Lebeda & Olson, 1994). Indeed, regions which do diverge are typically variable loop structures (Lebeda & Olson, 1994), including the respective sites for light chain cleavage which are all closely flanked by cysteine residues (Krieglstein *et al.*, 1994).

1.4.3 Neuron binding

Botulinal neurotoxins pass from the intestinal lumen to the lymphatic or vascular systems, possibly by a pinocytotic mechanism (Bonventre, 1979). The toxins are then considered to target cell surface receptors of peripheral cholinergic motor neurons which innervate the musculature.

Experiments to define the receptive molecule(s) to which the neurotoxins bind have suggested a number of possibilities. Early reports indicated that several neurotoxin serotypes would preferentially bind to presynaptic membranes in synaptosomal fractions

and could be inhibited by specific ganglioside molecules found naturally in neuronal membranes (Kitamura *et al.*, 1980). Type A neurotoxin interacts most strongly with ganglioside G_{T1b} and with synaptosomes in a temperature dependent manner (Kitamura *et al.*, 1980). The region of toxin binding was tentatively attributed to the extreme C-terminus of the heavy chain (Shone *et al.*, 1985). More recent evidence suggests that at least two receptor types may mediate toxin adherence to presynaptic terminals. One class is thought to be more prevalent, and of low affinity, whilst a second has a higher affinity, is less numerous and may be 'productive' in the intoxicating process (Daniels-Holgate & Dolly, 1996). The identity of a human, high affinity receptor for the type A neurotoxin has not been determined, although a 140 kDa presynaptic membrane glycoprotein with toxin affinity has been isolated from the Torpedo electric organ (Blasi *et al.*, 1992). However, multiple affinity receptors have been reported for the type B neurotoxin (Evans *et al.*, 1986), including gangliosides G_{T1b}, G_{D1a} and a protease-sensitive acceptor, possibly synaptotagmin (Ogasawara *et al.*, 1991; Nishiki *et al.*, 1994).

Tetanus neurotoxin, which binds to all peripheral and central neurons (Montecucco, 1986), may also bind to proteinaceous and ganglioside receptors (Halpern & Loftus, 1993). The ganglioside receptors bind to the 34 C-terminal amino acids of tetanus neurotoxin in a region which is conserved in other clostridial neurotoxins (Shapiro *et al.*, 1997). Thus, it seems that the C -terminus (Hc fragment) of the clostridial neurotoxins have a significant role in cell targeting and receptor docking, a prerequisite to endocytotic internalisation of the neurotoxin, and that a unique combination of ganglioside and protein receptors may be required for each toxin.

It is intriguing that synaptotagmin may form a neurotoxin receptor due to its involvement in Ca²⁺ dependant neurotransmitter release. This protein may be participating in both exocytosis and endocytosis at the presynaptic membrane (Geppert *et al.*, 1994)

as it is present on synaptic vesicles and also binds to the clathrin adaptor protein complex AP-2 (Chapman *et al.*, 1996).

1.4.4 Internalisation

Clostridial neurotoxins are thought to enter neuronal cells by an energy dependent receptor-mediated endocytotic mechanism, possibly involving clathrin-coated vesicles (Black & Dolly, 1986a,b). Having entered the neuron, the neurotoxins have been located to small transport vesicles which are then sorted (Black & Dolly, 1986b). Differences in the sorting process is believed to account for the dissimilar pathologies exhibited by tetanus and botulinal toxins. In contrast to botulinal toxins, vesicles containing tetanus toxin assume a retrograde mode of transport which directs it across synaptic clefts until the toxin has ascended to the Renshaw cells of the spinal cord. In contrast, botulinal toxin-containing vesicles may be recycled locally. (Ahnert-Hilger & Bigalke, 1995). The disparity in vesicle transport is dependent on the neurotoxin heavy chain. However, whilst it is presently unclear as to whether specific vesicles are selected or if the heavy chain directs vesicle transport is unknown, the N-terminus of the heavy chain fragment (Hn) is known to interact with lipid membranes (Shone *et al.*, 1987).

1.4.5 Vesicle membrane translocation

The precise mechanism by which the neurotoxic moiety of clostridial neurotoxins escapes from vesicles has not been determined. Strong evidence implicates the 50 kDa Hn fragment in this process (Boquet & Duflot, 1982; Mochida *et al.*, 1989; Hoch *et al.*, 1895; Montal *et al.*, 1992; Blaustein *et al.*, 1987). This domain forms channels in vesicle membranes which are stimulated by a reduction in pH of the vesicle lumen. The channel forming activity may be concomitant with pH-induced conformational changes and is

purported to involve a transmembrane domain composed of 4 amphipathic α -helices (Montal *et al.*, 1992; Lebeda & Olson, 1995). It is not clear whether a channel is formed *per se*, or if the membrane integrated proteins have a chaperone role, however, it is likely that multiple Hn fragments may aggregate to translocate the pharmacologically active entity (Donovan & Middlebrook, 1986).

1.4.6 Intracellular activity

Studies which by passed the neurotoxin binding and internalisation steps demonstrated that separation of the disulphide bridge connecting the light chain and heavy chain was necessary for intoxication (Stecher *et al.*, 1989), and that light chain alone was responsible for inhibiting exocytosis (Bittner *et al.*, 1989; Dolly *et al.*, 1990). The requirement of both chains in the *Aplysia* model appeared to be a characteristic of that particular non-vertebrate system (Maisey *et al.*, 1988).

Deletion mapping of the tetanus and botulinum light chain domains demonstrated that 8 amino acid N-terminal deletions or 65 and 32 C-terminal deletions (respectively) were still toxic, and a common motif remained which was characteristic of metalloproteases (Kurazono *et al.*, 1992). In agreement with these deductions an affinity site for Zn^{2+} was mapped to the putative metal binding site (Wright *et al.*, 1992) which complemented earlier findings that clostridial neurotoxins could be inactivated by divalent metal ion chelators (Bhattacharyya & Sugiyama, 1989). Thus, it appeared that the pharmacologically active component of clostridial neurotoxins could be an enzyme, which shared limited similarity with other Zn^{2+} -dependent metalloendopeptidases (Schiavo *et al.*, 1992).

Attempts by Schiavo and colleagues to discover a substrate for the clostridial neurotoxins in the subcellular fractions of neuronal cells identified a 13 kDa integral membrane

protein, termed vesicle associated membrane protein (VAMP) (also called synaptobrevin). Found on small synaptic vesicles (Montecucco & Schiavo, 1993), VAMP participates in the docking of these vesicles through an interaction with a presynaptic membrane protein called syntaxin (Sollner *et al.*, 1993). Vamp is proteolytically cleaved by tetanus neurotoxin and serotype B, D, F and G botulinal neurotoxins (Niemann *et al.*, 1994). It is interesting to note that only tetanus and type B neurotoxins have a common cleavage site in VAMP. Type A and E neurotoxins do not cleave VAMP at all, but are specific to another component of the synaptic vesicle docking apparatus, SNAP-25 (Schiavo *et al.*, 1993; Blasi *et al.*, 1993). Type C neurotoxin cleaves another member of the vesicle snare complex, syntaxin (Blasi *et al.*, 1993).

In retrospect, the distinct substrates of the neurotoxins complement earlier reports that particular neurotoxins exhibit 'quantitative and qualitative differences in their effects on neuromuscular transmission' (Sellin *et al.*, 1983). However, the specific activities of toxins involving proteinaceous components of the synaptic vesicle docking complex do not necessarily agree with other reported activities involving the activation of a transglutaminase (Facchiano *et al.*, 1993). This theory suggests that tetanus toxin is able to activate transglutaminase, which is then able to cross-link synapsin-vesicle complexes to actin filaments and prevent exocytosis. Current conjecture, however, suggests that tetanus toxin light chain is unable to reach a suitable concentration *in vivo* to activate transglutaminase (Ahnert-Hilger & Bigalke, 1995).

1.4.7 Functions of the non-toxic/haemagglutinin components

Although a great deal of data has been generated to describe the structure-function attributes of the 150 kDa 7 S neurotoxin, and to some extent the role of the 300 kDa 12

S complex in evading the hostile acidic and proteolytic environment of the stomach, little attention has been directed towards the haemagglutinin components of the larger 16 S and 19 S complex. Initial studies observed that a higher proportion of 7 S toxin was absorbed in the lymph when larger 16 S complexes were used instead of 12 S complex (Sugii *et al.*, 1977). This phenomenon was attributed to the higher stability afforded by the haemagglutinin components. However more recent evidence suggests that the haemagglutinin components have an active role in specific binding to the epithelium of the small intestine (Fujinaga *et al.*, 1997). The binding of the haemagglutinin components is thought to be with undefined glycolipids or glycoproteins which consequently mediate efficient absorption. This evidence could explain why 12 S toxins such as type E and F are comparatively less toxic by the oral route (Fujinaga *et al.*, 1997).

1.5 Genetics of neurotoxin production

1.5.1 Location of neurotoxin genes

One definition of virulence factors are those components of an organism whose loss specifically impairs pathogenicity but not viability (cited in Johnson, 1997). It is therefore, of no great surprise to find virulence factors associated with various mobile genetic elements, as the relative ease with which they are acquired and lost may offer a selective advantage. An association between mobile/extrachromosomal elements and neurotoxicity has been investigated in all known serotypes of *C. botulinum* and *C. tetani*. Not only are the neurotoxins dispensable, but in some cases show traits of genetic instability.

A link between phage and neurotoxicity was first demonstrated in a neurotoxin minus strain of *C. botulinum* type C, which could be converted back to a neurotoxic phenotype

by exposure to a lysate from a toxin producing strain (Inoue & Iida, 1970; Eklund *et al.*, 1971). Phage conversion was then demonstrated for *C. botulinum* type D, and, like type C, was considered to be pseudolysogenic (Eklund *et al.*, 1972). Strains cured of either toxigenic phage were then shown to be converted to neurotoxicity by either the 'type C' or 'type D' phages, indicating that the original host of either phage was a common clostridial strain (Eklund & Poysky, 1974). Phages encoding type C or type D neurotoxins are thought to be composed of a heterologous assemblage (Fujii *et al.*, 1988; Sunagawa & Inoue, 1991). Even though other serotypes of *C. botulinum* contained phage (Inoue & Iida, 1968) no other examples of phage-mediated toxin conversion were unequivocally demonstrated.

The production of a neurotoxin by *C. tetani* was correlated with the presence of a 75 kb plasmid. Furthermore, the spontaneous loss of which resulted in a toxin minus phenotype (Laird *et al.*, 1980). Oligonucleotide probes designed from the tetanus neurotoxin N-terminal amino acid sequence were then used to map the structural gene to a 16.5 kb *EcoRI* restriction fragment of this plasmid (Finn *et al.*, 1984). A similar scenario exists in *C. botulinum* type G strains, where a 114 kb plasmid is lost in non-toxigenic derivative strains (Eklund *et al.*, 1988). The genetic association between plasmid and neurotoxin complex were subsequently unequivocally demonstrated by Zhou *et al* (1995).

An extrachromosomal location for neurotoxin gene serotypes A, B, E and F has not been conclusively established. Type A strains possess both phages (Kinouchi *et al.*, 1981) and plasmids (Scott & Duncan, 1978; Strom *et al.*, 1984; Weickert *et al.*, 1986; Ferreira *et al.*, 1987), however, their presence has not been correlated with neurotoxicity.

PFGE studies of a type A *C. botulinum* strain were able to demonstrate that the neurotoxin gene was chromosomally located, but its loss was consistently the result of

the deletion of 70 kb (Lin & Johnson, 1995), specifically associated with the 'experimental' insertion of the conjugative transposon Tn916 (Johnson *et al.*, 1997). As the authors point out, an induced deletion of this nature may insinuate the presence of an unstable mobile genetic element, such as a prophage, insertion sequence or transposon.

Variations of a type B neurotoxin are produced by both proteolytic and nonproteolytic strains of clostridia which are not closely related (Collins *et al.*, 1994). This indicates that some type of horizontal gene transfer has occurred, but the genetic location of type B neurotoxins has not been established. Likewise, type F neurotoxins are found in more than one species of clostridia including proteolytic type F strains which can also be found in combination with either type A or type B neurotoxins (Gemenez & Ciccarelli, 1970; Hatheway & McCroskey, 1989), as well as non-proteolytic F strains and strains closely resembling *C. baratii* (Hall *et al.*, 1985). Again, there is no direct evidence that confirms the existence of a vehicle for this interspecies transfer.

Serotype E neurotoxin is produced by two distinct non-proteolytic clostridia, one resembles the clostridial type strain *C. butyricum* (McCroskey *et al.*, 1986; Collins *et al.*, 1994) whilst the more common strain resembles group II *C. botulinum* strains which produce type B and type F neurotoxins. Some evidence suggests that neurotoxic strains of *C. butyricum* contain a plasmid encoding the type E neurotoxin (Hauser *et al.*, 1992). However, the evidence obtained was indirect and is certainly less convincing compared to *C. tetani* and *C. botulinum* type G.

A separate study implicated a common prophage in toxigenic type E organisms, irrespective of species. Under the conditions used, the prophage could not directly infect non-toxigenic strains or hybridise with a type E neurotoxin-specific probe. Interspecies transfer was possible via the use of a second non-toxigenic helper strain (Zhou *et al.*,

1993). The precise genetic context of type E neurotoxin genes as well as those of other group I and II neurotoxin serotypes remains unclear.

1.5.2 Neurotoxin gene sequences

The first data obtained for botulinal neurotoxins was derived from N-terminal sequencing of the light chain, heavy chain and full length toxin of serotypes A, B and E (Schmidt *et al.*, 1984; Schmidt *et al.*, 1985; Sathyamoorthy & DasGupta, 1985). This data provided evidence that the neurotoxin was synthesised as a single polypeptide chain which was cleaved approximately one third from the N-terminus to yield anterior light chain and posterior heavy chain fragments. Furthermore the sequence data provided direct evidence that the botulinal neurotoxins A, B and E formed a homologous group which were related to tetanus neurotoxin (Robinson & Hash, 1982).

N-terminal sequence data was then to prove a decisive factor in the cloning and subsequent sequencing of the tetanus (Fairweather & Lyness, 1986; Eisel *et al.*, 1986), type A (Thompson *et al.*, 1990; Binz *et al.*, 1990), type C (Hauser *et al.*, 1990) type D (Binz *et al.*, 1990) and type E neurotoxins (Poulet *et al.*, 1992) whilst the remaining representatives were cloned using hybridisation and PCR strategies based on published neurotoxin serotypes (Elmore *et al.*, 1995).

Comparative alignment of the clostridial neurotoxin sequences revealed that each toxin was composed of a single ORF of approximately the same size (1251-1315 amino acids). Toxin genes had related light chains linked to related heavy chains coupled by conserved cysteine residues typical of the AB class of toxins. Each neurotoxin has a conserved zinc binding motif (HELxHxxHxLYG) centrally located in the light chain domain consistent with a metalloprotease activity. The primary sequences show blocks of conserved residue

groups occurring throughout the protein interspersed with variable regions thought to form loops in the native protein (Lebeda & Olson, 1994).

Variable regions are more conspicuous in the light chain domain and in the Hc fragment of the heavy chain consistent with the altered specificities of the neurotoxins substrates and ecto-receptors. Among the 153 conserved amino acids, only 4 amino acids corresponding to the zinc binding motif were conserved in the thermolysin sequence suggests that even though regions of predicted secondary structure were conserved, the evolutionary link predicted with the light chain was relatively distant (Lebeda & Olson, 1994). Regions of the high conservation appeared to correlate with four domains predicted to form transmembrane helices (Lebeda & Olson, 1995).

1.5.3 Genetic location of components from the neurotoxin complex

The progenitor neurotoxin complexes from a number of botulinal serotypes are known to contain proteins with haemagglutinin activity, whilst the 150 kDa neurotoxin or 300 kDa complex do not. The haemagglutinin components of some neurotoxin complexes are immunologically cross reactive (e.g., cross reaction is seen between types A and B or types C and D) suggesting that in addition to the neurotoxins, the haemagglutinin proteins may also be related (Somers & DasGupta, 1991). Moreover, in some instances the production of both neurotoxin and haemagglutinin were found to be associated with bacteriophage conversion in type C and D strains of *C. botulinum* (Oguma *et al.*, 1976).

It has subsequently been shown that the type C neurotoxin gene is closely associated with at least one haemagglutinin component of the type C neurotoxin complex and the genes are divergently transcribed (Tsuzuki *et al.*, 1990). Further sequence analysis has detected a third gene whose N-terminal sequence and size correlate with a 'non-toxic-nonhaemagglutinin' protein thought to form the 300 kDa complex with type C

neurotoxin gene (Tsuzuki *et al.*, 1992). Inspection of the region 5' to the neurotoxin gene of other serotypes suggests that the C-terminus of a similar gene may also reside next to other botulinal neurotoxins (Thompson *et al.*, 1990; Whelan *et al.*, 1992) but is not present in front of the tetanus neurotoxin gene (Fairweather & Lyness, 1986). Consequently, some evidence exists to postulate that components of some neurotoxin complexes may occur in close association with the respective neurotoxin gene such as in operons or regulons.

1.6 Detection of *C. botulinum*.

Botulism poses a serious risk to both producers and consumers of food. Current methods of detecting and diagnosing this condition include demonstrating the presence of vegetative/spores of neuropathogenic clostridia in the food samples, through selective culture, and via detection of neurotoxin using the toxin neutralisation/mouse bioassay, a slow, expensive and ethically unacceptable test. To improve this situation, advances in our knowledge of the toxins have allowed alternative strategies to detect either the organism or toxin directly. Two main approaches using either immunological or DNA hybridisation strategies have been used.

1.6.1 Immunological based detection methods

Early attempts to detect *C. botulinum* neurotoxins were principally based on immunological techniques, including radioimmunoassays, electroimmunodiffusion, ELISAs and immunodiffusion assays (Ferreira *et al.*, 1981). However, unless monoclonal antibodies were used, cross reactivity between neurotoxins or their progenitor complex components could contribute to confusion. To improve on the sensitivity of the ELISA assay approach, modifications have been made including the use of an enzyme-linked

coagulation assay which claims a sensitivity of <10 pg/ml which is comparable with the mouse bioassay (Doellgast *et al.*, 1994). This methodology offers advantages over alternatives, including colony immunoblotting (Goodnough *et al.*, 1993) and the use of a fiber optic-based sensor (Ogert *et al.*, 1992), in that it is both fast and does not require specialised equipment.

1.6.2 DNA hybridisation detection methods

Hybridisation forms the basis of a number of techniques by which specific DNA sequences are detected and has become particularly useful since the delineation of neurotoxin gene sequences. Among these, the combination of PCR and Southern hybridisation is of particular utility and sensitivity. A number of approaches have involved amplification of a specific region common to all botulinal neurotoxin genes of interest, followed by hybridisation of the PCR product to serotype specific fragments (Fach *et al.*, 1995; Campbell *et al.*, 1993). Although this technique offers a high degree of certainty, it is slow.

A more sophisticated method employs the amplification of serotype-specific sized PCR products which can be quickly analysed by gel electrophoresis (Whelan, S.M. Ph.D thesis 1993; Ferreira & Hamdy, 1995). Although this approach is possibly the most efficient, it is unable to distinguish between functional and silent genes (Franciosa *et al.*, 1994) and may not detect a small percentage of subtype or hybrid genes which do not hybridise with probe DNA. An alternative method of detecting group I *C. botulinum* strains is based on hybridisation of a DNA fragment specific to these neurotoxigenic organisms (McKinney *et al.*, 1993). This method does not appear to offer a selective advantage over PCR methods and is not thought to hybridize to elements of the neurotoxin

complex. It must, however, be of scientific interest as it is likely to represent some part of a mobile genetic element associated with neurotoxin genes in these organisms.

The most reliable detection methods are likely to be a combination of PCR and ELISA strategies which allows rapid detection and evaluation of the presence of a toxin molecule.

1.7 Clinical applications of neurotoxins

1.7.1 Clinical applications of neurotoxins

Botulinal neurotoxins are generally considered to possess a specific activity, the cleavage of vesicle associated proteins within synapses at the neuromuscular junction, which inhibits neurotransmitter release and hence inactivates muscular contraction. This property was recognised as a potential therapeutic agent in the treatment of movement disorders where paralysis of a particular muscle was desirable (Scott, 1980). In this respect the LL type A progenitor neurotoxin complex has principally been used. Due to its relatively large size, it does not diffuse extensively, so the region of muscular paralysis does not spread beyond the point of injection. A number of treatments for over active muscle disorders have become established including Blepharospasm (spasm of the orbicularis oculi muscle), Hemifacial spasm (a disorder of the facial nerve resulting in involuntary movements of facial muscles), Torticollis (deformity of the neck caused by contraction of cervical muscles and sternocleidomastoid muscles) and cosmetic applications such as removing wrinkles (Montecucco *et al.*, 1996). However the therapy is not permanent and must be periodically repeated. An immunological response attributed to the haemagglutinin moieties of the type A progenitor toxin complex can occur which impedes its further use. This could be circumvented by the use of serotypes (Montecucco *et al.*, 1996).

1.8 Clostridial genetics

1.8.1 Transcriptional factors

Those elements of the transcriptional machinery of clostridia that have been characterised to date share common features with the equivalent components of other eubacteria. The RNA polymerase holoenzyme of *C. perfringens* (Garnier & Cole, 1988) and *C. acetobutylicum* (Pich & Bahl, 1991), a basic unit of transcription, has been found to contain α , β and β' components. In addition, a number of σ factors have also been detected. A protein analogous to the *E. coli* σ^{70} /*B. subtilis* σ^A , which recognises vegetative/house keeping gene promoters, was detected in *C. acetobutylicum* by immunological cross reactivity. A similar preparation has been isolated from exponentially growing *C. perfringens* (Rood & Cole, 1991) and used to characterise promoter elements *in vitro*. In common with *B. subtilis*, clostridia have the ability to sporulate. Indeed, many of the σ factors which participate in the sporulation process of *B. subtilis* have homologs in *C. acetobutylicum*, including, σ^E , σ^F , σ^G , σ^H , σ^K , SpoOA and SpoDII. This indicates that sporulation is a conserved process in the separate genera (Wilkinson *et al.*, 1995). Furthermore, some clostridial genes appear to be under sporulation-specific control (including CPE, an enterotoxin produced by type A strains of *C. perfringens* (Melville *et al.*, 1997)) and under σ^D control, a σ -factor generally involved in the control of gene expression during late exponential phase (Sauer *et al.*, 1995).

Nucleotide sequence analysis of regions of DNA contiguous with cloned genes in solventogenic clostridia has revealed the presence of a number of transcriptional factors. In *C. beijerinckii* NCIMB 8052 a homolog of the LysR family was found adjacent to a thiolase gene (T. Davis, unpublished data), whilst MalR, a homolog of the *E. coli* LacI

repressor, has been found associated with an operon putatively involved in maltose metabolism (Dr N. P. Minton, pers. comm.). In *C. acetobutylicum* ATCC 824, a TetR homolog was discovered next to another thiolase gene, whilst in *C. acetobutylicum* P262 homologs of a two component response-regulator and the LacI repressor have been discovered (Woods & Reid, 1995). A σ^{54} homolog was also identified in *Clostridium kluyveri* (Sohling & Gottschalk, 1996). This transcriptional factor, which has been associated with nitrogen assimilation, pilin formation, motility and pathogenicity in Gram-negative bacteria, has also been implicated in the control of gene expression in the levanase operon of *B. subtilis* (Klier *et al.*, 1992). Thus, the limited sequencing undertaken has identified a variety of transcriptional factors. This situation resembles *B. subtilis* where there is, in addition to promoter-specific σ -factors, a range of transcriptional activators and repressors which have homologs throughout the bacterial kingdom, but which do not necessarily control the expression of equivalent operons and regulons present in other organisms (Moran, 1993).

1.8.2 Transcriptional signals

In keeping with the discovery of homologs of RNA polymerase and σ^{70} (also known as σ^A) in clostridia, a number of promoters have been characterized *in vitro* which resemble the target of the RNA- σ^A holoenzyme complex in *E. coli* and *B. subtilis*. The studies which were conducted, using 14 separate mapped promoters and the RNA- σ^A holoenzyme from *C. perfringens*, demonstrated that the derived consensus [5'-TTGAcA-14-TG-1-TATAAT-3'] was similar to the vegetative promoters of *E. coli* and like *B. subtilis*, in that it frequently had a -10 sequence preceded by a TG motif (Rood & Cole, 1991). Subsequent cloning of the σ^A gene from *C. acetobutylicum* has demonstrated that the regions of σ^A which are predicted to interact with the vegetative promoter regions

exhibit significant similarity to the equivalent regions of *E. coli* and *B. subtilis* σ^A , further indicating that the consensus vegetative promoters are also common to the clostridia (Sauer *et al.*, 1995). However, there are some examples of clostridial promoters which are not recognised by *E. coli* (including the UV responsive promoters of *C. perfringens* (Garnier & Cole, 1988a), and the P2 promoter of the *sol* operon transcript from *C. acetobutylicum* ATCC 824 (Fischer *et al.*, 1993)) suggesting that transcriptional factors with altered specificities are likely to be required in clostridia.

1.8.3 Translational signals

In contrast to *E. coli*, Shine-Dalgarno (SD) sequences preceding most clostridial genes have in general, a high level of complementarity to the 3' terminus of 16S rRNA (reviewed by Young *et al.*, 1989). Indeed, this phenomenon appears to be shared with *B. subtilis*, a paradigm of Gram-positive bacteria. The relationship between a 'strong' SD interaction and the efficiency of translation has been determined and has shown that in contrast to the Gram-negative *E. coli*, 'significant' complementarity is necessary for efficient translation initiation in *B. subtilis* (Vellanoweth, 1993). In the majority of cases, the initiation codon of clostridial genes is AUG, but as in *B. subtilis*, UUG and GUG also occurs (Young *et al.*, 1989).

1.9 Genetic manipulation of clostridia

The ability to genetically manipulate micro-organisms, particularly to generate defined mutants and complement the disrupted genotype via *in vivo* recombinant DNA techniques, has been an important goal in the study of bacteria. Within the clostridia, these have been developed predominantly in pathogenic or solventogenic species as they respectively command medical and industrial attention. Recent advances, such as the

construction of gene knockouts in *C. perfringens* (Rood, 1997), demonstrate the potential for genetic manipulation in these organisms, but currently, most clostridial species remain relatively recalcitrant to these techniques. However, this developing field employs an array of genetic tools.

1.9.1 Genetic elements

1.9.1.1 Indigenous plasmids

Several studies have established the widespread nature of plasmids in various species and strains of the clostridia (Minton & Morris, 1981; Truffant & Sebald, 1983; Urano *et al.*, 1983; Squires *et al.*, 1984; Popoff & Truffant, 1985; Lee *et al.*, 1987; Kurose *et al.*, 1989; Brehm *et al.*, 1992). Although plasmids with a wide range of molecular weights occur, research has focused on small plasmids (<10 kb) as a prelude to shuttle vector construction. Small plasmids are commonly found to encode replicons of the rolling circle type (Gruss & Ehrlich, 1989). In the clostridia these include pCB101 (Brehm *et al.*, 1992) and pNB2 (Delver *et al.*, 1996). Other cryptic clostridial plasmids which are not extensively characterised have been used as components of shuttle vectors. However, they do not necessarily form stable replicons (Yoshino *et al.*, 1990).

Unlike the solventogenic clostridia, where the indigenous plasmids have no selectable traits, genetic elements derived from *C. perfringens* commonly encode antibiotic resistance determinants and are consequently the basis of genetic systems in this organism (Heefner *et al.*, 1984; Squires *et al.*, 1984). A plasmid which has been well characterized, and is not thought to replicate via a rolling circle mechanism, is pIP404 (Garnier & Cole, 1988c). Vectors incorporating this clostridial replicon are a popular choice, particularly in the natural host *C. perfringens* (Shimizu *et al.*, 1997), as a replication mechanism which does not generate single stranded DNA intermediates is

more likely to be stable (Janniere *et al.*, 1990). Another plasmid of clostridial origin which has had its entire nucleotide sequence determined is the *C. butyricum* NCIB 7423 plasmid pCB102. This does not appear to possess a replicon resembling any other known plasmid (Minton *et al.*, 1993) and has not been developed into a shuttle vector.

1.9.1.2 Heterologous plasmids

Many bacterial plasmids are inherently promiscuous in nature and frequently encode various elements which promote inter-species transfer (Willems *et al.*, 1993). This observation has promoted the concept that these plasmid replicons may possess a wide host range. Thus, the initial transfer of plasmids into the solventogenic clostridia utilised genetic elements which originated from other Gram-positive bacteria (Reysset *et al.*, 1988).

In general, plasmids which are not of clostridial origin do not seem to form the basis for stable clostridial genetic systems (Truffaut *et al.*, 1989; Lin & Blaschek, 1984). Although many heterologous plasmid replicons are able to persist in the clostridia, they may be inefficiently maintained due to the incomplete recognition of all elements required for replication (del Solar *et al.*, 1996). Indeed, an undefined host mutation has been the basis by which one heterologous plasmid has attained stable inheritance in a clostridial host (Azeddoug *et al.*, 1992). However, recombinant plasmids which are derived from the *E. faecalis* conjugative plasmid pAM β 1 (Leblanc & Lee, 1984) are reported to possess traits of segregational stability and structural stability in solventogenic clostridia. This may be partly attributed to the theta mechanism of replication which does not generate highly recombinogenic single stranded DNA replication intermediates (Bruand *et al.*, 1991). However, there is still a requirement for well characterised, stable vector systems for the clostridia, particularly for thermophilic species.

1.9.1.3 Conjugative plasmids

This class of genetic element is defined by their ability to replicate autonomously as an extrachromosomal element and transfer between bacteria. Indeed, the host range for conjugative plasmids is typically broad and is complimented by a promiscuous replicative apparatus. Consequently, representatives of this type of plasmid are widespread throughout bacteria and are implicated in the 'horizontal' transfer of genetic information.

Although conjugative plasmids such as pCW3, pIP401, pJIR25 and pJIR27 have been recently identified within the clostridia (Lyras & Rood, 1997), the utility of conjugative gene transfer to the clostridia was demonstrated with heterologous conjugative plasmids of Gram-positive origin, including the MLS^R pAM β 1, pIP501 and derivatives (Young *et al.*, 1993). In addition to self-transmissibility, these plasmids have several interesting properties. Firstly, transfer between bacteria is mediated via the transfer of a single DNA strand which is not a substrate for many host restriction systems. Secondly, some non-conjugative plasmids can be co-mobilised from the donor, to the recipient bacterium. They are also a source of useful broad host range antibiotic resistance genes and stable replicons. However, the drawbacks for use in gene transfer include their large size, a degree of instability due to internal insertion sequences/transposons, and the introduction of self-transmissible elements into pathogens. Gene transfer by conjugation has been largely superseded by electrotransfer techniques.

1.9.2 Transposons

1.9.2.1 Conjugative transposons

In addition to the tetracycline resistance determinants *tetA*(P) and *tetB*(P) which are found linked on the *C. perfringens* conjugative plasmids homologous to pCW3, two

further unlinked tetracycline resistance genes are found in *C. perfringens* strains. One of these cross-hybridised with *tetM* and flanking regions of Tn916 indicating that this resistance determinant may be associated with a similar conjugative transposon (Lyras & Rood, 1997). A transmissible conjugative transposon designated Tn5397 is also thought to harbour a tetracycline resistance determinant in some strains of *C. difficile* (Lyras & Rood 1997). Nucleotide sequence analysis and hybridisation studies indicate a mosaic relationship between Tn916 and Tn5397 where some segments are virtually identical and some are unrelated (cited in Lyras & Rood, 1997). Erythromycin resistance can also be located on the conjugative transposon Tn5398 in *C. difficile* strain 630.

Heterologous conjugative transposons have been effectively used to generate mutants in both pathogenic and solventogenic clostridia. Although insertions have been reported in organisms such as *C. perfringens* (Rood *et al.*, 1997), *C. acetobutylicum* (Woolley *et al.*, 1989) and *C. botulinum* (Lin & Johnson, 1991), the occurrence of multiple insertions reduces the attraction of this technique.

The low number of transposable elements that have been detected in the clostridia may reflect that the majority of this large genera are not pathogenic and, therefore, have not benefited by obtaining antibiotic resistance determinants in this fashion. Transposable elements may be widely disseminated which encode cryptic functions not including antibiotic resistance; a parallel situation has already been exposed for plasmids in nonpathogenic clostridia.

1.9.2.2 Non-conjugative Transposons

Chloramphenicol resistance in *C. perfringens* is either encoded on a transposon, within a conjugative plasmid (i.e. pIP401 or pJIR25), or chromosomally. In either case, it is encoded by distinct chloramphenicol acetyl-transferase (CAT) genes. A homolog of the

plasmid encoded CAT gene of *C. perfringens* is found in multiple copies in the *C. difficile* genome. The transposon encoding CAT in pIP401 has been sequenced. In addition to encoding members of the resolvase/invertase and integrase proteins, a Pre/Mob type gene is present, indicating the potential for interspecies transmission by cointegrate formation or co-mobilisation. Again, the discovery of this transposon is related to its association with a selective phenotype.

1.9.3 Insertion sequences

An insertion sequence *IS1151* has been found next to the epsilon-toxin gene and/or enterotoxin gene in *C. perfringens* when these genes are present on a plasmid (Lyras & Rood, 1997). The association of this class of element with virulence factors is not uncommon and would thus seem a feasible means to facilitate toxin gene transfer. It has been postulated that the 'chromosomally' located botulinum neurotoxin type A gene may reside on a mobile genetic element (Johnson, 1997).

1.10 Gene transfer techniques in clostridia

1.10.1 Natural transformation

A number of bacterial species are able to absorb DNA from the surrounding milieu and incorporate it into their own genome (Stewart & Carlson, 1986). This process is facilitated if the DNA is both homologous and genomic, however, such 'natural competence' has not been documented in a *Clostridium*.

1.10.2 Transformation by chemical permeabilisation

A technique developed for the transformation of a *Bacillus brevis* strain, which was recalcitrant to natural transformation procedures, involved stripping the surface protein

layer with an alkaline Tris solution, followed by PEG-induced permeabilisation (Takahashi *et al.*, 1983). With minor modifications, this process was successfully applied to a mesophilic *C. acetobutylicum* strain (Yoshino *et al.*, 1990) and a thermophilic *C. thermohydrosulfuricum* strain (Soutschek-Bauer *et al.*, 1985).

1.10.3 Protoplast transformation

This approach to transformation relies on the natural competence of cell wall-less bacteria and thus protocols were developed to both generate cell wall-less protoplasts and subsequently to regenerate vegetative cells following transformation. Broadly speaking, many clostridial species were converted to protoplasts. However, efficient regeneration to vegetative cells proved problematical. In some cases regeneration appeared to correlate with the autolytic activity of the host (Heefner *et al.*, 1984; Reysett *et al.*, 1987). The most efficient protoplast transformation frequencies (10^5 transformants/ μ g of DNA) were achieved using an autolytic mutant of *C. acetobutylicum* N1-4080 (designated N1 4081), in combination with the addition of an autolytic inhibitor choline (Reysett *et al.*, 1988). The application of this technique to non-mutant strains of other clostridia was less successful.

1.10.4 Conjugative transfer

Conjugative methodology has had a number of applications in gene transfer to and between clostridial species. Unsuccessful attempts to simulate interspecies transfer of *C. perfringens* tetracycline resistance plasmids to *C. acetobutylicum* was successfully repeated with MLS^{R} plasmids from other Gram-positive bacteria (Oultram & Young, 1985; Reysett & Sebald, 1985). Having established a mode of gene transfer, complementation experiments were devised where a cloned gene was incorporated into

the wild type conjugal plasmid pAM β 1 by homologous cointegrate formation (Oultram *et al.*, 1988). Transfer from *B. subtilis* to *C. acetobutylicum* by conjugation demonstrated the utility of this approach. However, the application of a simpler technique, electrotransformation, superseded conjugation.

The success of conjugation prompted attempts to mobilise recombinant plasmids with only the origin of transfer (*oriT*) to clostridia. (Williams *et al.*, 1990). Transfer of shuttle vectors with the RK2 *oriT* were mobilised from an *E. coli* strain harbouring an IncP group helper plasmid. Although this technique is more 'cumbersome' than electrotransformation, entry of plasmid DNA via conjugative pili is in single stranded form and has the advantage of evading many host restriction systems.

1.10.5 Electrotransformation

The study of pore formation in cell membranes began with eucaryotic systems in which the application of an externally applied high-intensity electric field could induce membrane breakdown and cellular lysis (for a more recent review see Weaver & Chizadzhiev, 1996). Modulation of the electric field strength and duration produced conditions in which re-sealable pores could be formed and through which molecules could be introduced. The basic electroporation technique has now become an established tool by which molecules can be introduced into living cells representing most phyla of biology, not least the bacteria. Here, electroporation allows the introduction of genetic elements in the absence of natural competence or bypasses less amenable procedures, such as conjugation and protoplast transformation. The late eighties saw a plethora of both Gram-negative and positive bacteria successfully transformed using this technique (Trevors *et al.*, 1992), including two clostridial species *C. acetobutylicum* (Oultram *et al.*, 1988) and *C. perfringens* (Allen & Blaschek, 1988). Since this time, the

number of clostridial species and strains compliant with this technique has increased and includes members of group I *C. botulinum* (Zhou & Johnson, 1993).

Electrotransformation is rapid, inexpensive and reproducible, and can yield a high transformation efficiency. However, the introduction of heterologous DNA can still be subject to challenge by host barriers, including non-specific and site-specific nuclease activities (Mermelstein & Papoutsakis, 1993; Chen *et al.*, 1996).

1.11 Pathogenic bacteria

Bacteria survive and multiply by utilising the compounds available in their surroundings. Those bacteria which find a niche in, or on tissues, to the detriment of a host, are considered pathogenic. The factors which contribute towards the colonisation of a specific host environment, but otherwise are not vital to the microorganism's survival, are considered virulence factors. *C. botulinum* would appear to be an unusual pathogen because, in general, it does not invade host cells or tissues but acts remotely through the action of a potent neurotoxin. However, a survey of features which contribute towards pathogenicity in other bacteria may form a useful basis to compare the 'virulence factor(s)' and strategy of *C. botulinum*.

1.11.1 Virulence determinants

1.11.1.1 Toxins

Although toxic molecules derived from bacteria are relatively diverse, an initial classification can be imposed depending on whether the toxic molecule is a secreted protein (exotoxin), which are commonly produced by both Gram-positive and Gram-negative bacteria, or an endotoxin, which are lipopolysaccharides specifically derived from Gram-negative cell walls.

In this brief survey, exotoxins will be considered because, like clostridial neurotoxins, they are all proteinaceous molecules. Although isolated exotoxins can often elicit most symptoms of the respective disease, additional factors, such as the secretory apparatus, adherence factors and regulatory genetic circuits, are required to potentiate the action of the toxin (Finlay & Falkow, 1997).

Clostridial neurotoxins are examples of A-B toxins. This class of exotoxin is defined by the presence of a toxic entity (A), frequently an enzyme that modifies a molecule within the host, in combination with a second protein molecule(s) (B), which comprises a host specific binding domain which promotes delivery of (A) to a specific location of the host. Thus, exotoxins are combinations of enzymes and binding domains whose particular specificities directly contribute to the unique pathogenesis of an individual disease. In addition to clostridial neurotoxins, examples of A-B toxins include diphtheria and cholera toxins, whose A subunits possess ADP-ribosylating activity with distinct substrate specificities (London, 1992; Spangler, 1992), and shiga toxin, whose A subunit has a rRNase activity (Tesh & O'Brien, 1991).

A second class of unrelated enzymic toxins which modify strategic eucaryotic protein messengers include the *C. difficile* A and B toxin monoglucosyltransferases, *C. botulinum* exoenzyme C3 ADP-ribosyltransferase and *E. coli* cytotoxic necrotising factor type 1. These toxins modify members of the Rho family which mediate actin polymerisation (Aktories *et al.*, 1997) and consist of a single polypeptide chain.

A number of toxins target the host cell plasma membrane. Those which insert into the membrane to form pores include *E. coli* hemolysin, a member of the RTX hemolysins found in some Gram-negative pathogens (Finlay & Falkow, 1997), and lysteriolysin O, produced by *Listeria monocytogenes*, an example of a sulfhydryl-activated cytolysin found in Gram-positive pathogens (Farber & Peterkin, 1991). Plasma membranes are

also targeted via enzymic degradation. Phospholipase C is a major lethal factor produced by *C. perfringens* which has a number of roles in disease, particularly as a cytolytic factor (Titball, 1993) and is related to a number of other Gram-positive phospholipases.

A group of related toxins, that support their own secretion, are found in a range of Gram-negative pathogens and are represented by IgA proteases characterised from *Neisseria* spp. and *Haemophilus influenzae* (Finlay & Falkow, 1997). However their precise roles in pathogenesis remain largely undefined.

Unlike many of the exotoxins mentioned, two types of toxins, the heat stable toxins and superantigens, do not possess a catalytic function but mimic complex molecular interactions of the host. Heat stable toxins are similar to guanylin, a host activator of intestinal guanylate cyclase (Currie *et al.*, 1992), whilst superantigens, produced mainly by staphylococci, dramatically increase the indiscriminate interaction between antigen presenting cells and T cells, ultimately causing the elevated release of cytokines into the bloodstream where they induce fever and shock.

1.11.1.2 Adherence factors

Pathogens such as *C. botulinum*, *Staphylococcus aureus* and to some extent *C. perfringens* produce toxin(s) in food substances which interact with humans after ingestion. In contrast, the majority of pathogenic bacteria interact with or invade host cells and consequently are equipped with a variety of non-toxic adherence and invasion factors.

Adherence to specific host receptors is mediated by bacterial adhesins which commonly fall into two classes; fimbrial or afimbrial adhesins (Finlay & Falkow, 1997). Fimbrial adhesin (alternatively known as pili) genes are frequently heterogeneous and occur in

operons that may also encode protein chaperones which mediate fimbrial assembly (Hultgren *et al.*, 1991). The host ligand targeted by the fimbrial adhesin consequently determines tissue specificity during host colonisation. Afimbrial adhesins interact more closely with host cells and, in some cases, contain host binding motifs which mimic host factor molecules (Sandros *et al.*, 1993). Both Gram-positive and Gram-negative pathogens frequently possess adhesins which bind to the extracellular matrix (Patti *et al.*, 1994).

1.11.1.3 Invasive factors

A variety of pathogenic bacteria persist by entering a host cell and exploiting an intracellular niche. Not all bacteria share the same strategy, for example, *Salmonella*, *Mycobacterium* and *Legionella* gain access to intracellular vacuoles and remain there by altering the natural course of vacuole destruction, whilst *Listeria monocytogenes* and *Shigella flexneri* destroy the vacuole and occupy the cytoplasm (Finlay & Falkow, 1997). However, each micro-organism requires an invasive strategy to initially cross the host cell membrane. These mechanisms are typically complex and frequently involve interaction with host cell messengers. *Yersinia* spp. possess an outer membrane adherence factor, 'invasin', which binds to a host receptor. This interaction promotes bacterial engulfment in the host membrane and internalisation as a vacuole (Isberg & Leong, 1990). When other bacteria, such as *Salmonella typhimurium*, contact a host cell, a modulation of the host cytoskeleton is initiated, resulting in membrane ruffling and macropinocytosis of the bacterium (Francis *et al.*, 1993). In this instance, cytoskeletal rearrangement results from a bacterial factor interacting with a member of the host's Ras superfamily, Cdc42, one of many such factors involved in actin regulation (Nobes &

Hall, 1995; Chen *et al.*, 1996). Parallel mechanisms of host entry are also used by other Gram-negative intracellular pathogens (Finlay & Falkow, 1997).

Clearly, the methods by which different pathogens enter and survive in host cell environment depend on specialised virulence factors. Whatever the role a protein has in pathogenesis, e.g., toxin, adhesin, secretory protein, intracellular survival, immune evasion or host signal interaction, the regulation of these factors is of immense importance.

1.11.1.4 Regulation of virulence factors

A major form of protein regulation is exerted at the level of transcription. Bacteria have a number of factors which associate with RNA polymerase during the initiation of transcription. The success of transcription initiation depends largely upon the cumulative effect of the promoter/operator elements present and prevalence of transcriptional factors which interact with them. Genes which are vital to basic survival, with a 'house keeping' function, are controlled in this manner. Evidence suggests that specific virulence determinants from some organisms are to some extent also modulated by factors similar to those which participate in 'housekeeping' regulation. For example, homologs of the AraC family participate in virulence regulation activation in *Yersinia* spp. (Skurnik & Toivanen, 1992), *Shigella flexneri* (Sakai *et al.*, 1988) and *Vibrio cholerae* (Higgins & DiRita, 1996). Likewise, members of the LysR regulatory family have been identified in co-ordinate regulation of some virulence genes in *Salmonella* spp. (Caldwell & Gulig, 1991).

In contrast to AraC and LysR type regulatory factors which modulate transcription by binding to operators which facilitate RNA polymerase binding, some virulence gene promoters have activators which replace the sigma entity of the RNA polymerase and

thus exert promoter specificity. However genes which are regulated by 'alternative sigma factors' are not all involved in pathogenicity *per se*, but in general could be classified as factors which promote survival. For instance, 'extracytoplasmic function' factors (ECF) are a group of small sigma type proteins which appear to participate in the regulation of specific operons (Lonetto *et al.*, 1994) in both Gram-positive and Gram-negative bacteria. Members of this group have been implicated in controlling pathogenic determinants (Schurr *et al.*, 1995) as well as a variety of stress responses (Huang *et al.*, 1997). The specific mechanism by which ECF factors are activated or transcribed is as yet unclear. However, transcriptional control can in some cases be subject to 'regulatory cascades' of environmental response messengers and intermediate transcriptional factors, capable of interacting with disparate genetic loci and thus co-ordinating an environmental response. Such mechanisms often involve two component regulators which were initially found to be responsive to environmental conditions, such as osmolarity or nitrogen, phosphate and carbon source concentrations (Russo & Silhavy, 1993). These systems are also known to promote a large number of virulence responses (Dziejman & Mekalanos, 1995) and have been characterised in Gram-positive bacteria, including *C. perfringens* (Shimizu *et al.*, 1997).

Recent evidence has demonstrated that a distinct system of intercellular communication exists between Gram-negative bacteria. This system is comprised of a diffusible messenger (N-acyl homoserine lactone (N-AHL), a N-AHL synthase and [N-AHL] dependent transcriptional activator (Salmond *et al.*, 1995). The prevalence of such 'quorum sensing' systems appears wide spread among Gram-negative bacteria and is intimately linked with the production of virulence determinants by some Gram-negative pathogens (Salmond *et al.*, 1995). Although Gram-positive bacteria do not seem to possess N-AHL-dependent quorum sensing potential, a number of cell density dependent

systems have been described which appear to be mediated by pheromone peptides (Kleerebezem *et al.*, 1997), and include the virulence response of *S. aureus*.

There is tentative evidence to suggest that an intercellular signal molecule participates in the regulation of some *C. perfringens* virulence genes, but it is not established whether this is a cell density sensing mechanism (Shimizu *et al.*, 1997).

In conclusion, as regulation by promoter activity can be influenced at a number of levels through the action of regulatory cascades, the potential for modulation via intercellular and intracellular signals is immense and reflects the complex and coordinated nature of bacterial responses.

1.11.1.5 Pathogenicity islands ~ inheritance and dissemination of virulence

‘Pathogenicity island’ is a term coined to describe a defined region of a genome in which are encoded a number of virulence determinants often with associated functions (Lee, 1996). Analysis of these features has revealed that they are commonly associated with some type of mobile genetic element which is either extrachromosomal, for example a phage or plasmid (e.g., in *Yersinia* and *Shigella* human pathogens, (Rosqvist *et al.*, 1995)) or a similar structure which has integrated into the pathogens genome (e.g., retronphages in *E. coli* enteric pathogens (Inouye *et al.*, 1991)). It is unlikely that inheritance of genetic information in this manner is entirely associated with virulence as large cryptic plasmids are widespread, but rather that virulence factors are one facet of the bacterial kingdom’s array of tools which are employed to gain a selective advantage by colonising a new environment. For a non-pathogenic example, essential genes required for solventogenesis (a metabolic response to unfavourable environmental conditions) are plasmid borne in some clostridial species (Girbal & Soucaille, 1998). The association of multiple virulence factors on a single ‘island’ would be logical when they

are interdependent i.e., operons containing toxin genes and respective secretion apparatus, or two synergistic toxins. The acquisition of pathogenicity islands may therefore be a common route by which non-pathogenic bacteria become virulent.

It is known that regulation of certain genes encoded within pathogenicity islands are influenced by regulatory factors which are also found in the pathogenicity island, which may in turn be controlled by chromosomally encoded mechanisms (Johnston *et al.*, 1996). Furthermore, cross-talk between elements of distinct pathogenicity islands may also contribute to virulence regulation (Morschhauser *et al.*, 1994).

1.12 Research aims

This project will attempt to determine the genetic basis of neurotoxin complex production in *C. botulinum* NCTC 2916, with a specific focus on the genetic context of the silent type B neurotoxin gene in this strain. To complement these studies, a novel plasmid-based gene transfer system will be sought for use in *C. botulinum* (group I), providing a vehicle by which the regulation of neurotoxin complex production can be studied *in vivo*.

CHAPTER TWO

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals and kits.

Amersham International, Buckinghamshire, U.K.

Hybond N+ Nylon membrane, megaprime DNA labelling kit.

Life Technologies Ltd. (GibcoBRL), Paisley, U.K.

Agarose, ammonium persulphate, caesium chloride, calcium chloride, DNA kilobase ladder, ethylenediaminetetraacetic acid (EDTA), ethidium bromide, formamide (37 % v/v), formaldehyde, glycine, Lambda DNA *Hind*III fragments, magnesium chloride, mineral oil, phenol, phenol:chloroform:isoamyl alcohol 25:24:1, pre-prepared dialysis tubing, sodium acetate, sodium hydroxide, sucrose, tetra-methylethylenediamine (TEMED), T4 Polynucleotide kinase/buffer.

Biorad laboratories, Watford, U.K.

Acrylamide, methylene-bis acrylamide.

Boehringer Mannheim (BCL), East Sussex, U.K.

AMV reverse transcriptase, deoxynucleotide triphosphates, calf intestinal alkaline phosphatase, Klenow DNA polymerase I fragment, restriction endonucleases/buffers, and T4 DNA ligase.

CP Laboratories, Herts, U.K.

RNase A.

Du Pont (U.K.) Ltd. Hertfordshire, U.K.

α -[³²P]dATP (aqueous), γ -[³²P]rATP (aqueous) and α -[³⁵S]dATP (aqueous).

iGi (Immunogen International Ltd) Tyne & Wear, U.K.

Deoxynucleotide triphosphates.

Original TA Cloning® Kit.

Northumbria Biochemicals Ltd.

Restriction endonucleases/buffers.

New England Biolabs (U.K.), Ltd. Herts, U.K.

Restriction endonucleases/buffers.

Perkin Elmer, Cheshire, U.K.

Amplitaq® DNA polymerase, 10x PCR buffer, 25 mM MgCl₂ solution and 10 mM dNTP kit.

Pharmacia Biotech Ltd. Herts, U.K.

Sequenase™ version II.

Sigma Aldrich Company Ltd, Dorset, U.K.

Absolute ethanol, ampicillin, dimethylformamide, erythromycin, guanidinium thiocyanate, isoamyl alcohol, kanamycin, lysozyme (chicken egg white), polyethylene glycol (PEG 8000), polyoxyethylene 20 cetyl ether (Brij 58), β-isopropyl-thio-β-galactosidase (IPTG), propan-2-ol, sodium dodecyl sulphate (SDS) and 5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside (X-gal).

Stratagene Ltd, Cambridge, U.K.

RNase block.

2.1.2 Nucleic Acids.

Oligonucleotide primers were synthesised on an Applied Biosystems 380A DNA synthesiser and were kindly supplied by Mr R. Hartwell and Dr M. Duffield, (Division of Structural Sciences, CAMR, U.K).

VECTOR	SOURCE OF REPLICON	SELECTION	FEATURES	REFERENCE
pMTL20	Col E 1	Amp ^R	Cloning vector	Chambers <i>et al.</i> , 1988
pMTL22	Col E 1	Amp ^R	Cloning vector	Chambers <i>et al.</i> , 1988
pMTL20E	Col E 1	Amp ^R /Erm ^R	Replicon probe vector	Minton <i>et al.</i> , 1993
pCB4	Col E 1/pCB102	Amp ^R /Erm ^R	pCB102 replicon clone	Minton <i>et al.</i> , 1993
pMTL500E	Col E 1/pAMβ1	Amp ^R /Erm ^R	Broad-host range shuttle vector	Minton <i>et al.</i> , 1993
pMTL500F	Col E 1/pAMβ1	Amp ^R /Erm ^R	Broad-host range expression vector	Brehm <i>et al.</i> , 1987
pFD	Col E 1	Amp ^R	Ferredoxin clone	Graves <i>et al.</i> , 1985
pGK12	pWV01	Erm ^R /Cm ^R	Broad-host range lactococcal vector	Kok <i>et al.</i> , 1984
pXcm I	Col E 1	Amp ^R /Kan ^R	PCR cloning vector	Cha <i>et al.</i> , 1993
pCR2.1	Col E 1	Amp ^R /Kan ^R	PCR cloning vector	Invitrogen Inc

Table 2.1 Plasmids used in this study

2.1.3 Bacterial strains.

STRAIN	GENOTYPE	REFERENCE
<i>E. coli</i> DH5α™	F' <i>supE44</i> Δ(<i>lacZYA-argF</i>),U169 (φ80 <i>lacZ</i> Δ <i>M15</i>) <i>hsdR17</i> (r _k ⁻ m _k ⁺) <i>recA1</i> <i>endA1</i> <i>deoR</i> <i>gyrA96</i> λ <i>thi-1</i> <i>relA1</i> <i>phoA</i> .	Woodcock <i>et al.</i> , 1989.
<i>E. coli</i> InvαF'	F' <i>endA1</i> <i>recA1</i> <i>hsdR17</i> (r _k ⁻ ,m _k ⁺) <i>supE44</i> <i>thi-1</i> <i>gyrA96</i> <i>relA1</i> φ80 <i>lacZ</i> αΔ <i>M15</i> Δ(<i>lacZYA-argF</i>) U169.	Invitrogen ®.
<i>E. coli</i> GM2163	F' <i>ara-14</i> <i>leuB6</i> <i>thi-1</i> <i>fhuA31</i> <i>lacY1</i> <i>tsx-78</i> <i>galK2</i> <i>galT22</i> <i>supE44</i> <i>hisG4</i> <i>rpsL136</i> (Str ^r) <i>xyl-5</i> <i>mtl-1</i> <i>dam13::Tn9</i> (Cam ^r) <i>dcm-6</i> <i>mcrB1</i> <i>hsdR2</i> (r _k ⁻ ,m _k ⁺) <i>mcrA</i> .	Palmer & Marinas, 1994.
<i>E. coli</i> HB101	F' Δ(<i>gpt-proA</i>)62 <i>leuB6</i> <i>supE44</i> <i>ara-14</i> <i>galK2</i> <i>lacY1</i> Δ(<i>mcrC-mrr</i>) <i>rpsL20</i> (Str ^r) <i>xyl-5</i> . <i>mtl-1</i> <i>recA13</i> .	Boyer & Roulland-Dussoix, 1969.
<i>E. coli</i> TG1	F' <i>traD36</i> <i>lacI</i> ^a Δ(<i>lacZ</i>) <i>M15</i> <i>proA</i> ⁺ <i>B</i> ⁺ / <i>supE</i> Δ(<i>hsdM-mcrB</i>)5 (r _k ⁻ m _k ⁻ <i>mcrB</i>) <i>thi</i> Δ(<i>lac-proAB</i>).	Gibson, 1984
<i>Clostridium beijerinckii</i> NCIMB 8052	wild type <i>C. beijerinckii</i> (Solventogenic).	NCIMB
<i>Clostridium botulinum</i> type A NCTC 2916	wild type <i>C. botulinum</i> encoding type A and silent B neurotoxin genes	NCTC
<i>Clostridium botulinum</i> type A NCTC 7272 / ATCC 19397	wild type <i>C. botulinum</i> encoding type A neurotoxin gene.	NCTC
<i>Clostridium botulinum</i> type A ATCC 3502	wild type <i>C. botulinum</i> encoding type A neurotoxin gene.	ATCC
<i>Clostridium botulinum</i> type B 'Danish'	wild type <i>C. botulinum</i> encoding type B neurotoxin gene.	Danish Serum Institute
<i>Clostridium botulinum</i> type B ATCC 25765 (17B)	wild type <i>C. botulinum</i> encoding type B neurotoxin gene.	ATCC

Table 2.1 Bacterial strains

The list of strains is accompanied with a genotype reference where possible, however, the clostridial strains are unpublished wild type isolates which can be traced to the respective culture collection.

2.2 METHODS

2.2.1 Storage and maintenance of bacterial strains

Strains of *E. coli* were stored at -70°C using the Microbank™ system (Prolab Diagnostics). Stocks were revived by plating onto M9 minimal media (Appendix II) supplemented with the appropriate metabolic requirements and incubated at 37°C for 16 h. Cells were maintained at 4°C on LB agar plates for 4 weeks.

Culture and manipulation of anaerobic bacteria was conducted when ever possible in a Mk 3 anaerobic work station (Don Whitely Scientific Ltd, West Yorks, U.K.) The atmosphere, N₂:CO₂:H₂, was maintained at ratios of 87:9:4 respectively and at a temperature of 37 °C. Clostridial strains were inoculated into 20 ml of CMC broth (Appendix II) incubated for 24 h and stored. Single colonies were obtained by spreading 10 µl of storage broth on to a blood-agar plate (Appendix II) and incubated for 16-24 h.

2.2.2 Purification of DNA

2.2.2.1 Small scale plasmid isolation from *E. coli*

Small quantities of plasmid DNA were routinely isolated by a modification to the 'Rapid Boiling' method of Holmes and Quigley (1981). A single colony was used to inoculate 2 ml of L-broth (Appendix II), supplemented with the appropriate antibiotic. Cultures were incubated aerobically for 5-16 h at 37°C with shaking (225 r.p.m). Cells were harvested by centrifugation (6000 x g, 3 min) and the supernatant completely removed. The cell pellet was resuspended in 100 µl of STET buffer (Appendix I) containing 10 mg/ml lysozyme and placed on ice for 2 min. Cells were lysed in a boiling water bath for 40 sec. and the cellular debris pelleted by centrifugation (13000 x g, 15 min). The pellet was removed and plasmid DNA precipitated at -70°C for 20 min by the addition of 160 µl of propan-2-ol. Plasmid DNA was recovered by centrifugation (13000 x g, 10 min, 4°C) and the supernatant completely removed. The pellet was air dried and resuspended in 50 µl of 1xTE buffer (Appendix I). RNA contamination was removed by adding 1 µl of RNase A (10 mg/ml) per sample to the STET buffer.

2.2.2.2 Medium scale plasmid isolation from *E. coli*.

Superior quality plasmid DNA was isolated using the 'Plasmid Midi Kit' (Qiagen Ltd. Surrey, U.K.) according to the manufactures instructions. The system is based on the alkaline lysis procedure of Birnboim & Doly (1979).

2.2.2.3 Large scale plasmid isolation

To obtain large quantities of plasmid DNA, a procedure was adapted from the method of Clewell and Helinski (1969). Typically 1 l of broth was inoculated with a single fresh colony and incubated at 37°C for 16 h with shaking (225 r.p.m). Those transformed with the pWVO1 plasmid replicon (Kok *et al.*, 1984) were inoculated with 2 ml of a late exponential culture. Cells were harvested by centrifugation (6000 x g for 20 min, 4 °C) and the supernatant completely removed. The cell pellet was resuspended in 5 ml of Tris-sucrose solution (Appendix I) and placed on ice for 2 min. Lysozyme was added to 2 ml of 0.25 M EDTA and 1 ml of 0.025 M Tris-HCl (pH 8.0) at a final concentration of 10 mg/ml. This solution was added to the cells, mixed by gentle inversion and incubated on ice for a further 5 min. Lysis was induced by the addition of 8 ml of Brij/DOC solution (Appendix I). Cellular debris was pelleted by centrifugation (30,000 x g, 40 min, 4°C). Plasmid DNA was isolated from the cleared lysate by isopycnic centrifugation (Section 2.2.2.4).

2.2.2.4 Isopycnic centrifugation

Plasmid DNA from crude lysates was purified from chromosomal DNA, RNA and protein contaminants by isopycnic centrifugation in a continuous gradient of caesium chloride (Sambrook *et al.*, 1989). Crude lysates containing 1.25 g/ml caesium chloride and 80 µg/ml ethidium bromide were centrifuged at 160,000 x g for 48 h in a T12-70 rotor at 19°C. The plasmid band was removed by aspiration and ethidium bromide extracted from the sample three times with caesium chloride-saturated isopropanol, before dialysis against 5 l volumes of 1x TE buffer (4°C) at 1 h, 4 h and 16 h intervals (Sambrook *et al.*, 1989). DNA samples were stored at - 20°C.

2.2.2.5 Small scale bacteriophage isolation

Single plaques were inoculated into 3 ml of 2x YT broth and incubated with shaking (225 r.p.m) for 16 h at 37°C. Cells were pelleted by centrifugation (13,000 x g for 5 min) and the supernatant added to a clean tube. To this, 250 µl of precipitation was added (20 % (w/v) PEG MW 8000, 2.5 M NaCl in sterile H₂O) and the mixture incubated at room temperature for 15 min. Precipitate was collected by centrifugation (13,000 x g, 5 min) and the supernatant completely removed by aspiration. The pellet was resuspended in 150 µl of 1x TE and extracted with 50 µl of phenol:100 x TE buffer solution (1:0.4). The mixture was centrifuged (13,000 x g, 2 min) and 140 µl of the aqueous layer removed to a fresh tube containing 280 µl of chilled absolute ethanol and 14 µl of 3 M sodium acetate. DNA was precipitated at -70°C for 20 min and harvested by centrifugation (13,000 x g, 2 min). The pellet was washed with 70 % (v/v) ethanol and air dried before being dissolved in 40 µl of 1x TE.

Double stranded bacteriophage DNA could be isolated from the cell pellet by the method described in section 2.2.2.1.

2.2.2.6 Large scale bacteriophage DNA isolation

A single plaque and an uninfected *E. coli* strain (TG1) were individually inoculated into 10 ml of 2xYT broth prior to 16 h incubation at 37°C. The 10 ml of uninfected culture was inoculated into 1 l of 2x YT broth pre-warmed to 37°C and incubated until mid-exponential growth phase (O.D₆₀₀ = 0.6). The infected 10 ml of culture was centrifuged (13000 x g, 10 min) and the supernatant added to the 1 l culture of exponentially growing *E. coli*. Incubation was continued for a further 4 h before a cleared lysate was obtained by the method described in section 2.2.2.3 and purified replicative form (RF) bacteriophage obtained by the method described in section 2.2.2.4.

2.2.2.7 Isolation of DNA from clostridial strains.

Plasmids were purified from clostridial species by a variation of the alkaline lysis method (Birnboim & Doly, 1979) as described by Ish-Horowicz and Burke (1981). Cells from 1.5 ml of culture, grown to late exponential phase (O.D₆₀₀ = 0.8-1.2) were pelleted by centrifugation (13,000 x g, 5 min). The supernatant was completely

removed, the cells gently resuspended in 100 µl of GTEL buffer (Appendix I) and the mixture incubated at 37°C for 15 min. A further 200 µl of lysis solution (Appendix I) was added and after brief mixing by vortexing, the lysing cells were placed on ice for exactly 5 min. Cellular debris was precipitated by the addition of 150 µl of chilled 3 M potassium acetate solution (pH 4.8) and incubated on ice for 30 min. The precipitate was pelleted by centrifugation (13,000 x g, 20 min, 4°C). The supernatant fraction was extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and nucleic acids precipitated by the addition of 2.5 vol. of absolute ethanol and 0.1 vol. of 3 M sodium acetate solution (pH 4.8). The pellet was collected by centrifugation (13,000 x g, 20 min 4°C), washed briefly with 70 % (v/v) ethanol and then resuspended in 100 µl of 1xTE buffer. RNase A was added to a final concentration of 10 µg/ml and the mixture was incubated for 30 min at 37°C. After further phenol:chloroform extraction, the plasmid DNA pellet was resuspended in 1xTE buffer. To obtain larger quantities of plasmid, the ratio of cell culture volumes to volume of buffers was kept constant. In addition, an isopycnic centrifugation step (see section 2.2.2.4) was introduced after RNase treatment to remove contaminating chromosomal DNA and proteins.

2.2.2.8 Phenol Extraction

DNA and RNA samples were routinely purified by phenol extraction. An equal volume of phenol equilibrated with 100 x TE buffer, or a phenol:chloroform:isoamyl alcohol (25:24:1) solution was added to the sample and mixed by shaking for 20 sec. After 1 min the sample was again shaken for 20 sec and centrifuged (13,000 x g, 5 min). The upper aqueous layer was promptly removed for further manipulations.

2.2.2.9 Column Purification

DNA samples were routinely purified and concentrated using the Wizard clean up column, Promega Ltd. Southampton, U.K. or the Qiagen® column systems, Qiagen Ltd, Surrey, U.K and used according to the manufacturers instructions.

2.2.2.10 Ethanol precipitation

Nucleic acids were precipitated from solution by the addition of 2.5 vol. of chilled absolute ethanol and 0.1 vol. of 3 M sodium acetate (pH 4.8). Precipitation was enhanced by placing the samples at -70°C for 20 min. Nucleic acids were recovered by centrifugation (13,000 x g, 20 min 4°C), the supernatant completely removed, and the pellet washed in 70 % (v/v) ethanol before being air dried. The sample was finally resuspended in 1 x TE buffer and stored at -20 °C.

2.2.2.11 Analysis of DNA.

DNA was quantified at 260 nm using quartz cuvettes with a 1 cm path length and 1 ml volume (Sambrook *et al.*, 1989). An O.D₂₆₀ of 1.0 equated to concentrations of 50 µg/ml of double stranded DNA, 40 µg/ml of single stranded DNA and 20 µg/ml of synthetic oligonucleotide respectively.

Contamination of DNA with protein or phenol was estimated by determining the ratio for a sample at both O.D₂₆₀ and O.D₂₈₀. A ratio (O.D_{260/280}) below 1.8 indicated contamination. The structural integrity of nucleic acid was also judged visually on agarose gels.

2.2.3 Transformation of bacteria.

2.2.3.1 Preparation of competent *E. coli*.

Competent cells were prepared by a modification of the method of Cohen (1972). A single colony was inoculated into 50 ml of L-broth and grown with shaking (225 r.p.m) at 37°C to an O.D₆₀₀ of 0.8. The culture was cooled on ice for 10 min and harvested by centrifugation (6000 x g, 10 min). After removal of the supernatant, the cells resuspended in 30 ml of ice cold 0.1 M MgCl₂ solution, the cells were pelleted (6000 x g, 5 min) and finally resuspended in 2.5 ml of ice cold 0.1 M CaCl₂ solution. Competency was induced by incubation on ice for a minimum of 2 h. Cells could be stored at +4 °C for 24 h without loss of transformation efficiency.

2.2.3.2 Transformation of competent *E. coli* cells with plasmid DNA.

Cells were transformed by the addition of 0.1 volumes of DNA (typically 1-10 µl) to an aliquot of competent cells, the mixture was placed on ice for 20 min and heated at 42 °C for 2 min before replacing on ice for 20 min. A period of expression was included when using bacteriocidal selection before plating the transformed cells. 300 µl of L-broth was added to the cells and incubated at 37 °C for 20 min.

2.2.3.3 Transfection of competent *E. coli* with bacteriophage DNA.

Competent cells were prepared by the method presented in section 2.2.3.1 using a strain with a F' episome, commonly strain TG1. Phage DNA was mixed with competent cells and placed on ice for 20 min prior to a 'heat shock' step of 42°C for 2 min, the cell suspension was then placed on ice for 5 min and added to 3 ml of molten H-top agar (Appendix II) at 45°C with 200 µl of an exponentially growing culture of *E. coli* (O.D₆₀₀=0.6-0.8). The mixture was poured on to a 2x YT agar plate (Appendix II) and allowed to set before incubation at 37°C for 16 h.

2.2.3.4 Transformation of *C. beijerinckii* strains.

C. beijerinckii NCIMB 8052 was transformed using a variation of the method described by Oultram *et al.*, (1988). A single colony was used to inoculate 10 ml of YTG media (Appendix II). This was serially diluted (10^{-1} - 10^{-4}) to give 5 cultures, which were grown for 16 h at 37°C. The culture which had not reached stationary phase was used to inoculate a 100 ml volume of YTG broth and grown to an O.D₆₀₀ = 0.8. The culture was cooled on ice, harvested by centrifugation (6000 x g, 10 min) and washed once in ice cold S.E.B electroporation buffer (Appendix I) before resuspension in 5 ml of the same buffer. Electroporation was conducted using 0.3 ml of cell suspension in a 0.2 cm electroporation cuvettes, applying a 1.25 kv pulse, 200 Ω, 25 µF in a Bio-Rad gene Pulser™. The mixture was then incubated at 37°C in 3 ml of YTG broth for 3 h and then collected by centrifugation (13,000 x g, 5 min). The cells were resuspended in 100 µl of YTG broth and plated onto dry YTG plates supplemented with erythromycin (20 µg/ml). Manipulations were performed in an anaerobic environment whenever feasible

using an Anaerobic Workstation Mark III (Don Whitley Scientific Ltd, West Yorkshire, U.K) .

2.2.4 Determination of plasmid stability

2.2.4.1 Structural stability in *C. botulinum*/*C. beijerinckii*.

Plasmid DNA was isolated from the transformed organism and re-transformed into an *E. coli* strain. Larger quantities of plasmid were isolated from *E. coli* and then subjected to restriction enzyme analysis, using an equivalent control whenever applicable.

2.2.4.2 Segregational stability in *C. beijerinckii*.

The stable inheritance of plasmids by daughter cells was estimated by a method described by Swinfield (1987). A single transformed colony was inoculated into 10 ml of YTG broth supplemented with erythromycin (20 µg/ml). The culture was incubated for 8 h where upon an aliquot of 50 µl was used to inoculate a further 10 ml of YTG. This culture was also grown for 8 h and the sub-culturing step repeated, finally after a third 8 h of growth, serial dilutions were spread on to non-selective YTG agar plates. The resulting colonies were replica plated onto both selective (erythromycin, 20 µg/ml) and non-selective YTG plates. The segregational stability of a plasmid was deduced by the percentage of colonies still retaining an antibiotic resistant phenotype.

2.2.5 Electrophoresis of Nucleic Acids

2.2.5.1 Horizontal slab gel electrophoresis of DNA.

DNA was analysed by horizontal electrophoresis using 0.7-2.0 % (w/v) agarose gels. These were cast in portable casting trays (5 x 8 cm, 11 x14 cm and 20 x 21cm) by melting agarose in the appropriate buffer (Appendix I) using a microwave oven. Loading wells were created by immersing combs in the molten agarose. Samples were loaded using DNA tracking dye (Appendix I) and separated by electrophoresis at a voltage of between 25 - 100 V/cm. DNA was stained using ethidium bromide and illuminated using short wave u.v. (254 nm). Photographic records were taken with a Polaroid MP4 land camera with a red filter, using Polaroid 667 film. Visualisation of DNA for preparative purposes was done using long wave length u.v. (366 nm).

2.2.5.2 Horizontal slab gel electrophoresis of RNA.

RNA molecules were analysed on agarose gels according to the method of Goda & Minton (1995). Apparatus was pre-soaked in aqueous 1 % (w/v) SDS solution and rinsed with sterile H₂O prior to use. Electrophoresis was performed according to section 2.2.5.1 except that gels were composed of 1.2 % (w/v) agarose, and that before pouring, guanidinium thiocyanate was added to the molten agarose at a final concentration of 20 mM. Agarose solutions and electrophoresis buffers comprised of double autoclaved 1x TBE buffer (Appendix I). Prior to loading, 10 µl of RNA (section 2.2.10.2) was added to 5 µl of formaldehyde (38 % v/v) and 10 µl formamide and boiled for 2 min, before the addition of 3 µl of 10x RNA tracking dye (Appendix I). Electrophoresis was performed at 15 V/cm.

2.2.5.3 Recovery of DNA from agarose gels.

DNA fragments were recovered from an agarose matrix by an adaptation to the 'electrophoretic elution method' of McDonell (1977). Narrow slices of agarose were inserted within ¼ " washed dialysis tubing with a small volume of 1x TBE buffer (pH 8.0). DNA was electroeluted out of the gel slice into the dialysis bag by electrophoresis at 100 V/cm. The solution in the dialysis bag was transferred to a clean tube and the DNA purified according to section 2.2.2.9.

2.2.5.4 Denaturing polyacrylamide gel electrophoresis.

Single stranded linear nucleic acids were separated through denaturing polyacrylamide gels based on a system described by Sanger (1980). Polyacrylamide was cast and run between two glass plates (20 x 50 cm), which were thoroughly cleaned and rinsed with absolute ethanol. The 'notch plate' was coated with a gel repellent (2 % (v/v) dimethyl-dichlorosilane in 1,1,1-trichloroethane in absolute ethanol whilst the 'back' plate was coated with a gel adherent (75 µl of γ -(methyl-acryloxy)-propyltrimethoxysilane and 750 µl of 10 % (v/v) glacial acetic acid in 25 ml absolute ethanol). The dried plates were separated by a narrow 1 mm spacer and sealed. A gradient gel was made with 40 ml of 0.5 x TBE gel mix and 7 ml of 5x TBE gel mix (gel mixes A and B respectively, Appendix I). Polymerisation was initiated by the addition of 5 µl/ml 25 % (w/v)

ammonium sulphate and 1 µl/ml TEMED. Approximately 4 ml of mix B and 7 ml of mix A were poured between the glass plates. The remaining volume of mix A was added in such a manner as to create a regular vertical gradient with mix B. Shark tooth or square tooth combs were inserted to create sample wells and the polymerisation allowed to continue for a further hour. Gel rigs were run with upper and lower reservoirs of 1x TBE running buffer at 1500 V/ 25 mA. The gel sandwich was parted and the plate with the adhered gel was fixed in 10 % (v/v) glacial acetic acid containing 10 % (v/v) methanol for 15 min. DNA sequencing gels were dried at 80°C before autoradiography.

2.2.6 Modification of DNA with enzymes

2.2.6.1 Restriction endonucleases.

Double stranded DNA was cleaved with restriction endonucleases in the specified buffers provided by the enzyme's manufacturers. No more than 1/10 of the final digest volume consisted of the restriction endonuclease's storage solution. Digests were incubated at the temperature suggested by the manufacturer. Digests using more than one enzyme were conducted in a mutually compatible buffer. Restriction enzymes were heat inactivated at 70°C for 10 min before further manipulation of the DNA.

2.2.6.2 Klenow Fragment.

The large fragment of *E. coli* DNA polymerase I (Klenow fragment) was used for creating 'blunt' ends on ds DNA from 3' recessed termini. The reaction was conducted at 37°C in 1x TM buffer (Appendix I) using 5 units of enzyme and dNTPs at a final concentration of 25 µM. DNA fragments were then purified according to section 2.2.2.8 and 2.2.2.9.

2.2.6.3 Calf Intestinal Alkaline Phosphatase.

Calf intestinal alkaline phosphatase (CIP) was used to remove the 5' phosphates from linear plasmid DNA to prevent subsequent re-ligation. 20 U of CIP were used in a reaction containing 1-10 µg of DNA in 1x phosphatase buffer (Appendix I). The mixture was incubated at 37°C for 20 min and the enzyme inactivated at 70°C for a further 10 min. Plasmid DNA was then purified according to either section 2.2.2.8 or 2.2.2.9.

2.2.6.4 T4 polynucleotide ligase.

Purified DNA fragments with compatible termini were covalently joined *in vitro* using T4 DNA ligase. Varying ratios of the candidate DNA fragments of approximately 50 ng/μl were incubated at 14°C for 4-16 h with 0.5 units of ligase in 1x ligase buffer (Appendix I).

2.2.7 Immobilisation of nucleic acids on Nylon membranes

2.2.7.1 Southern blotting.

Double stranded DNA was transferred from agarose gels to charged nylon membranes using a variation of the method of Southern (1975). DNA was first separated by electrophoresis and then partially depurinated by soaking the gel in 0.25 M HCl for 45 min with gentle agitation. After brief rinsing in sterile H₂O, the DNA in the gel was denatured by soaking in 1.5 M NaCl, 0.5 M NaOH for 45 min. After a further wash in sterile water, the alkaline solution was neutralised with 0.5 M Tris-HCl, 1.5 M NaCl (pH 7.5) for 30 min. Denatured DNA was then transferred to a positively charged nylon membrane (Hybond N⁺, Amersham International, Bucks, U.K.) via capillary blotting, using 400 ml of sterile 20 x SSC as transfer buffer (Appendix I) and Quick Draw™ blotting paper (Sigma, Poole, U.K.). Following a 4-16 h blotting period, the membrane was laid upon a pad of 3 MM Whatman filter paper pre-soaked in 0.4 M NaOH for 20 min to covalently fix the DNA to the membrane. The membrane was briefly washed in 5x SSC and either probed for DNA sequences or stored at 4°C.

2.2.7.2 Northern Blotting.

RNA gels were blotted by capillary action in sterile 20x SSC transfer buffer as previously described (Section 2.2.7.1), to positively charged nylon membranes. The RNA was then attached to the membrane covalently by heating at 80°C *in vacuo* for 2 h and probed for mRNA species.

2.2.7.3 *In-situ* colony hybridisation.

Individual plasmid clones were isolated from a genomic library by a variation of the method of Grunstein and Hogness (1975). *E. coli* transformants were replica plated on to Hybond N' nylon membrane which was in contact with L-agar (Appendix II)

membrane was carefully peeled off the agar and laid on several layers of 3 MM Whatman filter paper soaked in 0.4 M NaOH. Cell lysis and attachment of the plasmid DNA to the membrane occurred during the next 20 min, the membrane was washed briefly in 5x SSC and then immersed in GOOP pre-hybridisation solution (Appendix I) at 56°C. Residual cellular debris was reduced by replacing the pre-hybridisation solution after 1 h of rotation in a hybridisation oven.

2.2.8 Preparation of DNA Probes

2.2.8.1 Radio-labelling of DNA probes.

Double stranded DNA fragments to be labeled were isolated by gel electrophoresis in a low melting point agarose gel (Section 2.2.5.1) The required fragment (50-100 ng) was prepared and labeled with α -[³²P]dATP using the Megaprime™ kit (Amersham International, Bucks, U.K.) according to the manufacturers instructions. Briefly, a pool of random hexamer nucleotides were annealed to the DNA probe and extended with Klenow fragment in the presence of α -[³²P]dATP and BdNTP nucleotides. Radio-labeled probes were added directly to the pre-hybridised membrane without further purification.

2.2.8.2 Radio-labelling of DNA size markers.

DNA kilobase ladder or lambda DNA-*Hind*III fragments were radio-labeled by filling in their terminal recesses with α -[³²P]dATP. DNA polymerase 'Klenow fragment', dGTP, dCTP, dTTP and 1x TM buffer (Appendix I) were supplied by Boehringer Mannheim (BCL), East Sussex, U.K. The labeling reaction was incubated for 30 min at 37°C. A volume of 1 μ l was loaded per well of an agarose gel.

2.2.9 Hybridisation.

2.2.9.1 Hybridisation of probes to nucleic acids.

Membranes were prehybridised in 'GOOP' pre-hybridisation solution (Appendix I), typically for 15 min. Radio-labelled probe was added and the incubation continued at the appropriate hybridisation temperature for 16 h. Unbound probe was removed by washing 3 times using 2 x SSC, 0.1% (w/v) SDS. The membrane was then air dried, wrapped in Saran™ Wrap and auto-radiographed against Kodak X-OMAT AR film. To enable the repeated probing of membranes for different sequences, the previous probe

was removed using one of two methods; membranes containing DNA were immersed in stripping solution containing 0.4 M NaOH at 50°C for 30 min. This solution was replaced with strip neutralisation buffer (Appendix I) and gently agitated for a further 15 min. Finally the membrane was rinsed in 2x SSC and the efficiency of probe removal qualified by autoradiography. Alternatively, membranes containing immobilised RNA were stripped by immersion in a boiling solution of 1 % (w/v) SDS. This was continued until the SDS solution had cooled to 40°C, The membrane was then washed in 2x SSC and residual probe detected by autoradiography.

2.2.9.2 Autoradiography.

Kodak X-OMAT AR film was routinely used for the detection of ³⁵S and ³²P radioactive isotopes. Film was placed next to a dried sequencing plate or Nylon membrane in an X-ray film cassette. Calcium tungstate-phosphor intensifying screens were employed to improve sensitivity. The sealed cassette was stored at -80°C for 16-72 h before developing the film in a Kodak X-OMAT automatic developing machine.

2.2.10 Purification and Analysis of RNA.

2.2.10.1 Culture of strains.

Preparation of cells for harvesting RNA was adapted from the method of Henderson *et al.*, (1996). A fresh 20 ml volume of CMC broth (Appendix II) was inoculated with 1.5 ml of CMC storage culture of the appropriate clostridial strain and incubated at 37°C for 24 h. An aliquot of this culture, typically 1 ml, was used to inoculate 20 ml of USA II broth (Appendix II) supplemented with 1 % (w/v) glucose. This culture was incubated for 16 h and used as an inoculum for a further 80 ml of USA II broth containing 1% (w/v) glucose. Incubation was continued for a further 5 h.

2.2.10.2 Extraction of total RNA from *C. botulinum* strains.

During all procedures solutions, glassware, tubes and disposables were 'double autoclaved' at 121°C for 15 min to eliminate RNase contamination. Cells from 10 ml of *C. botulinum* culture were harvested by centrifugation (6000 x g, 5 min at 4 °C), the supernatant completely removed and the cell pellet was then resuspended in 500 µl of chilled sterile AE buffer (Appendix I). To this, 50 µl of 10 % (w/v) SDS was added and

then mixed thoroughly by vortexing. An equal volume of phenol, previously equilibrated with 1x AE buffer, was added and the cells mixed again by vortexing. The mixture was incubated in a 65°C water bath for 5 min and then frozen in a dry ice-ethanol water bath for 2 min. The sample was then centrifuged (13,000 x g, 4 min), the aqueous layer extracted with phenol chloroform (section 2.2.2.8) and then the RNA was precipitated. DNA was removed by resuspending the RNA pellet in 50 µl of 1x DNase buffer (Appendix I) containing 20 units of DNase I and incubation at 37°C for 1 h. The RNA was then extracted once more using phenol:chloroform and finally precipitated with ethanol. RNA was stored under the ethanol supernatant at -70°C until required.

2.2.10.3 Analysis by Primer Extension.

The 5' nucleotide of a mRNA species was determined by the method of Faulkner (1992). Approximately 30 nM of oligonucleotide primer was phosphorylated at the 5' terminus with γ -[³²P]rATP in the following reaction: 1 µl of primer solution was mixed with 5 µl of γ -[³²P]rATP, 3 µl of 10x kinase buffer (Appendix I), 2 µl of T4 polynucleotide kinase and 19 µl of H₂O. The mixture was incubated at 37°C for 1 h. The volume was subsequently made up to 130 µl with 1x TE, phenol extracted (section 2.2.2.8) and ethanol precipitated (section 2.2.2.10) with the inclusion of 1 µl of glycogen to aid formation and visualisation of the precipitated oligonucleotide pellet. End-labeled primer was then resuspended in 10 µl of 3x PEB buffer (Appendix I), to which 5 µl of total RNA was added and 15 µl of double autoclaved water. Primer was annealed to the target mRNA (Section 2.2.10.2) using a MJ-Research thermal cycler, programmed for an initial denaturing step at 80°C for 10 min followed by annealing for 3 h at 55°C. The oligonucleotide-mRNA hybrid was precipitated in 60 µl of chilled absolute ethanol, air dried and resuspended in a mixture 1x reverse transcriptase buffer (Appendix I), 1mM of each dNTP, actinomycin (25 µg/ml) and 1 µl of placental RNase inhibitor. The extension reaction was initiated by the addition of 9 U of AMV reverse transcriptase and incubated at 42°C for 1 h. Extension products were precipitated by the addition of 50 µl of chilled absolute ethanol and 2 µl of sterile 3 M sodium acetate and incubated at -20°C for 1h. The precipitate was collected by centrifugation (13,000 x g, 10 min) and air dried. The DNA was resuspended in denaturing running buffer (Appendix I), heated at 85°C for 2 min and then loaded along side control sequencing

reactions, generated from the same oligonucleotide and template onto a 6 % denaturing polyacrylamide gel (Section 2.2.5.4).

2.2.10.4 Analysis by Primer Extension-Blotting.

This method is identical to primer extension except that the oligonucleotide used was non-radioactive. After running the polyacrylamide gel, extension products were blotted on to Hybond N⁺ membrane using 20 x SSC as transfer buffer. The ss DNA was then fixed to the membrane with 0.4 M NaOH for 20 min, concomitantly hydrolysing the alkali labile RNA. A PCR fragment was generated which was complementary to the putative primer extension product, this probe was radioactively labeled (section 2.2.8.1) and used to probe for specific primer extension products (Section 2.2.9.1).

2.2.11 DNA Sequence Analysis

2.2.11.1 Primers.

Oligonucleotide primers of between 17-20 bases were designed from known DNA sequence and synthesised (section 2.1.2). Palindromes, inverted repeats and compatible ends were avoided whilst the G+C content was maximised to increase the annealing temperature. Oligonucleotide stock solutions were quantitatively determined (section 2.2.2.11) using a 1/200 fold dilution in H₂O. The concentration was calculated in µg/ml and adjusted to the working concentration of 30 µM/ml for ds plasmid sequencing and 2 µM/ml for ss M13 sequencing.

2.2.11.2 Plasmid template.

Double stranded DNA plasmid template was prepared using commercially available column based procedures (section 2.2.2.9). Between 1-3 µg was used for each sequence reaction.

2.2.11.3 M13 bacteriophage template.

Superior quality template was obtained by cloning DNA fragments into RF M13 bacteriophage vectors M13mp18 and M13mp19 (Messing *et al.*, 1986). Single stranded DNA template was purified according to the method described in section 2.2.2.5.

2.2.11.4 Manual sequencing reactions.

DNA was sequenced 'manually' using the chain terminating method described by Sanger (1977). Thermal denaturation was used to prepare the template for sequencing (Andersen *et al.*, 1992). Briefly, 8 µl of plasmid DNA and 1 µl of oligonucleotide primer solution were annealed by heating to 97°C for 5 min and cooled to the annealing temperature of 37°C. Immediately 1 µl of 10x TM buffer was added and annealing continued for a further 20 min. The primer was extended at room temperature with 5.5 µl of labeling mixture (Appendix I) for 5 min. The sample was then split into four equal fractions and individually mixed with 3.5 µl of a dideoxy-nucleotide termination mix (Appendix I) and incubated at 37°C for 10 min. The chain termination reaction was stopped with the addition of 4 µl of formamide stop dye (Appendix I) and either stored at -20°C or denatured for 2 min at 70°C before loading on to a polyacrylamide gel for electrophoretic separation (section 2.2.5.4). Ss M13 templates/primer solutions were denatured in a water bath at 80°C and allowed to anneal slowly in the cooling water bath reservoir, an appropriate volume of 10 x TM buffer was added and the sample was then sequenced as described for ds plasmid DNA.

2.2.11.5 Automated DNA sequencing reactions

Double stranded plasmid DNA was sequenced with an automated robotic catalyst using approximately 25 µg of plasmid DNA prepared using the Qiagen midi system (section 2.2.2.2) and 4.5 pmol of oligonucleotide primer. Sequence reactions were extended by the cycle sequencing method using *Amplitaq* FS dye terminator chemistry. Samples were separated and recorded using an ABI377 DNA sequencer.

2.2.11.6 Analysis of data.

DNA sequence data was collated and processed using the 'DNASTAR' program 'LASERGENE', DNASTAR Inc. London U.K.

2.2.12 Polymerase Chain Reaction (PCR)

2.2.12.1 Oligonucleotide primers.

Oligonucleotide primers were designed to achieve several objectives, to amplify DNA fragments with an elucidated DNA sequence, to amplify DNA fragments whose

sequence was predicted and to amplify known sequences, but in addition to engineer specific base changes in the primer sequence to create, for example, convenient restriction endonuclease sites. Whilst designing primer pairs, several factors were considered, complementarity between regions within a single primer and between two primers was avoided, especially at the 3' end. Mesophilic clostridia have a high A+T DNA base bias particularly in the intergenic regions, so, where possible, primers were designed with a high percentage G+C DNA base content, thus increasing the temperature and specificity of annealing. The length of primers was generally between 30-36 base pairs. This enhanced specificity, and helped overcome steric repulsion between the template and mutagenic primers.

2.2.12.2 Template preparation.

Purified chromosomal and plasmid DNA samples were diluted 1/1000 using sterile H₂O. Small scale preparations of genomic DNA were also obtained from single colonies using the Gene Releaser™ system (Cambio, Cambridge, U.K).

2.2.12.3 Reagents and conditions.

A thermostable DNA polymerase was used in the amplification reaction (Saiki *et al.*, 1988). Biotaq® (Bioline Ltd.) and Amplitaq® (Perkin Elmer) DNA polymerases were chosen as they possess a non-template directed polymerisation activity, often resulting in the addition of a dATP residue and consequently facilitating the cloning of the amplified fragment in 'T' vectors. Reaction volumes were 100 µl, comprising of 10 µl of template, 10 µl of each primer at 10x working concentration (0.2 µM), 70 µl of 'PCR reaction master mix', typically containing 2.5 units of polymerase, 10 µl of 10x PCR buffer (Appendix I) and dNTP's at a final concentration of 0.2 mM. A 50 mM MgCl₂ stock solution was added to give a final concentration of MgCl₂ between 1.5 - 3.0 mM.

PCR amplification was attempted using a number of thermal programs, typically an initial template denaturation step was conducted at 94°C for 1 min, followed by multiple cycles of three steps starting with a denaturation step of 94°C for 1 min, a primer-template annealing step at a temperature between 37°C and 60°C depending on the primers, and finally an incubation step at 72°C, where the polymerase extends the primer to complement the template strand. This cycle was repeated 25-30 times and

followed with a further extension step at 72°C to complete the synthesis of the amplified products. Products were analysed by agarose gel electrophoresis..

2.2.12.4 Cloning PCR products.

DNA fragments amplified by certain DNA polymerases commonly have a single non-template directed nucleotide, often a dATP residue polymerised to the 3' end (Hu, 1993). These fortuitous overhangs can be used to clone PCR fragments into complementary 'T' vectors. Two such vectors were used. The 'TA' Cloning® kit vector pCR2000 (Invitrogen BV, Leek, the Netherlands) was used according to the manufacturers instructions. An alternative 'T' vector, p*Xcm* I was kindly supplied by Dr S. Chandrasegaran (Cha *et al.*, 1993). Double stranded plasmid DNA was isolated and digested with the endonuclease *Xcm* I whose recognition sequence is 5'...CCANNNNN↓NNNNTGG...3'. Two of these sites flank a kanamycin resistance gene and possess dTTP residues at the appropriate position to create compatible ends for PCR products. Ligated PCR products were transformed into competent *E. coli*, selected on ampicillin agar plates and then replica plated onto kanamycin agar plates. Plasmid was isolated from transformants not possessing a kanamycin resistant phenotype and analysed for a cloned insert by restriction mapping.

CHAPTER 3

GENETIC CHARACTERISATION OF A SILENT TYPE B NEUROTOXIN GENE FROM *C. botulinum* NCTC 2916.

3.1 Introduction.

During the course of a previous project in this laboratory, a method of detecting clostridial neurotoxin genes was developed using PCR and neurotoxin serotype specific oligonucleotide primers. The amplified products were indicative of the type of neurotoxin gene present (Dr S. M. Whelan. Ph.D thesis, 1993). In general, the results could reproducibly determine the presence of a specific neurotoxin gene in a particular strain but did not amplify DNA fragments from any other organism tested. However, the type A strain of *C. botulinum* NCTC 2916 consistently produced results indicating the presence of both type A and type B neurotoxin genes. Successive isolation of clonal colonies and screening with serotype specific primers, including primers derived from the type B strain Danish, indicated the presence of an entire type B gene. In contrast, antibody neutralising tests indicated that only serotype A toxin was expressed in the strain, suggesting that the gene encoding the type B gene was silent (Dr S. M. Whelan, Ph.D thesis, 1993).

The concept of a clostridial strain containing two copies of neurotoxin genes is not new. Several examples of strains expressing dual neurotoxin combinations of serotypes A, B and F have been documented (section 1.3.2). However, in these cases, both neurotoxins were detectable. The aim of this study was, therefore, to genetically characterise the type B neurotoxin gene and determine the genetic basis of why the type B neurotoxin gene in *C. botulinum* NCTC 2916 was not expressed.

3.2 Strategy.

In order to manipulate genomic clones from *C. botulinum* in a rational and safe manner, a high resolution restriction map of the silent neurotoxin locus was required. To achieve this, a number of DNA probes likely to hybridise to the target gene were required. Fortunately, a representative serotype B neurotoxin gene had been sequenced (Whelan *et al.*, 1992) and was an obvious candidate as a source of DNA sequence data and probes. Having established a restriction map, specific DNA fragments were then targeted for cloning which, for safety reasons, were theoretically incapable of encoding a neurotoxic peptide.

3.3 Cloning of the neurotoxin locus.

3.3.1 Cloning of a DNA fragment 5' to the silent neurotoxin gene.

Data accrued from the type A neurotoxin and proteolytic type B neurotoxin DNA sequencing projects (Thompson *et al.*, 1990; Whelan *et al.*, 1992) were used to design two oligonucleotide primers. These were anticipated to selectively amplify a 622 bp DNA fragment encompassing the 3' extremity of an ORF preceding the silent neurotoxin gene, a short intergenic section and the 5' end of the silent neurotoxin gene itself (figure 3.1). The 5' oligonucleotide, designated B2676, was 22 bases in length and homologous with the extreme 5' sequence known for the proteolytic type B neurotoxin locus (Whelan *et al.*, 1992). In comparison to the equivalent region in the type A locus, nine base mis-matches occurred, six of which were at the 3' end. It was, therefore, unlikely that B2676 would anneal with the type A sequence. The oligonucleotide primer designated B2677 was homologous with the proteolytic type B neurotoxin coding sequence in a region where only 10 of 23 bases were in common with the type A neurotoxin. PCR amplification with B2676, B2677 and *C. botulinum* NCTC 2916 genomic DNA produced a DNA fragment of the expected size. The DNA was recovered from the agarose matrix by electroelution and purified by phenol extraction. The DNA fragment was then cloned into pCRII and designated pSB06.

3.3.2 Investigation of the 5' coding sequence.

The insert of pSB06 was sequenced with forward and reverse universal primers and the oligonucleotides B2676 and B2677 (figure 3.1). Comparison of the sequence data with the equivalent region of the published proteolytic type B sequence (Whelan *et al.*, 1992) revealed a fragment of identical length with 7 nucleotide substitutions, 6 of these occurred in the coding region of an upstream ORF resulting in 2 conserved amino acid changes and 4 non-conserved amino acid substitutions. A further nucleotide substitution occurred in the intergenic region. No obvious discrepancy could account for the proposed transcription or translational dysfunction of the neurotoxin gene. Both the coding sequence of the upstream ORF and the putative neurotoxin gene remained in frame, whilst the proposed ribosome binding site and initiation codon were homologous to the published sequence. A possibility existed that one or more of the nucleotide substitutions may have disrupted a promoter element, however, no data currently

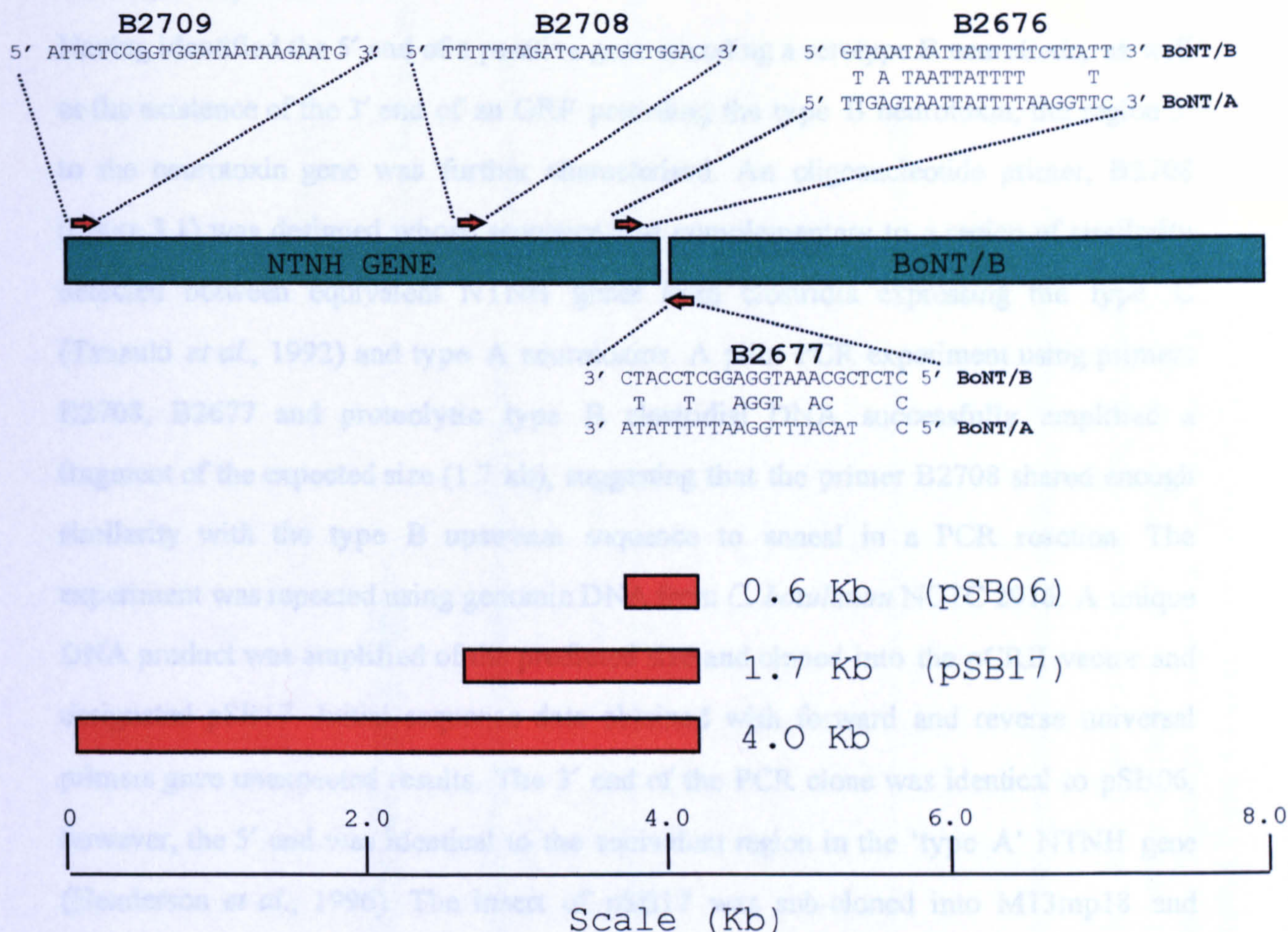


Figure 3.1 Selective PCR amplification from the silent type B neurotoxin locus.

Three separate DNA fragments were amplified from the type B neurotoxin locus in strain NCTC 2916 (tan boxes). Structural genes are represented by green boxes. Oligonucleotide primers used in the PCR are shown below their designated name and where pertinent, are compared to the equivalent region in the type A locus. Primers 2676 and 2677 were designed in regions where the type A and type B neurotoxin and NTN gene sequences varied, whilst primers B2708 and B2709 were designed where conserved motifs occurred in the type A and type C NTN genes

documents promoter(s) for this gene. Another consideration is that the nucleotide substitutions observed may have been artifacts of the amplification process (Tindall & Kunkel, 1988).

Having identified the 5' end of a putative gene encoding a serotype B neurotoxin, as well as the existence of the 3' end of an ORF preceding the type B neurotoxin, the region 5' to the neurotoxin gene was further characterised. An oligonucleotide primer, B2708 (figure 3.1) was designed whose sequence was complementary to a region of similarity detected between equivalent NTNH genes from clostridia expressing the type C (Tsuzuki *et al.*, 1992) and type A neurotoxins. A pilot PCR experiment using primers B2708, B2677 and proteolytic type B clostridial DNA successfully amplified a fragment of the expected size (1.7 kb), suggesting that the primer B2708 shared enough similarity with the type B upstream sequence to anneal in a PCR reaction. The experiment was repeated using genomic DNA from *C. botulinum* NCTC 2916. A unique DNA product was amplified of the predicted size and cloned into the pCRII vector and designated pSB17. Initial sequence data obtained with forward and reverse universal primers gave unexpected results. The 3' end of the PCR clone was identical to pSB06, however, the 5' end was identical to the equivalent region in the 'type A' NTNH gene (Henderson *et al.*, 1996). The insert of pSB17 was sub-cloned into M13mp18 and M13mp19 bacteriophage vectors using the flanking *EcoRI* restriction sites of the pCRII plasmid. Further sequence data obtained with primers designed to the type A sequence demonstrated that the majority of the insert of pSB17 was homologous to the type A NTNH gene and that the 5' end of pSB06 was very close to the end of this homologous tract (see figure 4.1).

One explanation for this observation could be that a chimeric PCR product (Bhavsar *et al.*, 1994) may have been formed between the type A and type B upstream genes in the PCR reaction, possibly due to inefficient primer annealing at the 5' end of the type B upstream gene or because B2708 efficiently annealed to both type A and B upstream genes.

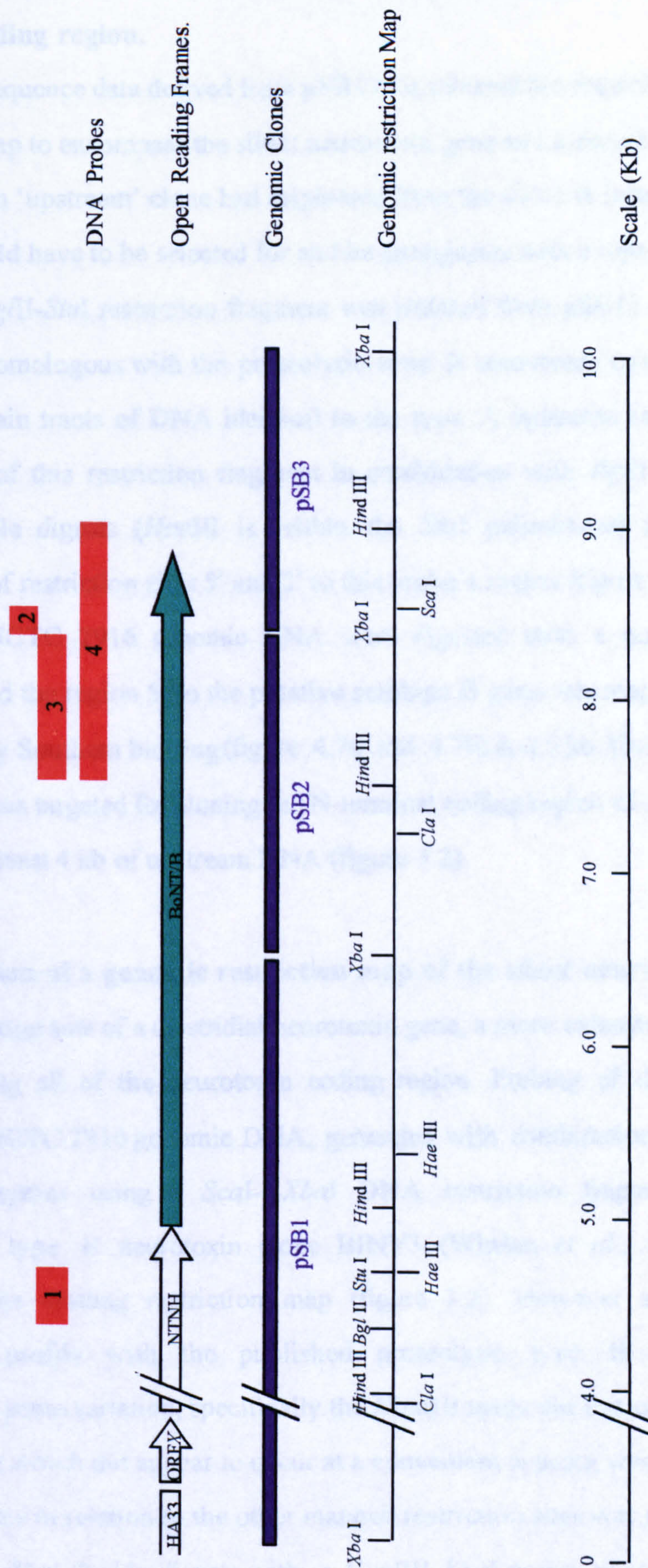


Figure 3.2 Schematic representation of the strategy used to clone genomic fragments of the type B neurotoxin locus

DNA probes (1-4) were used to generate a genomic restriction map, from which the three contiguous *Xba* I restriction fragments were targeted for genomic cloning. Probe 1 was a *Bgl* II-*Stu* I restriction fragment from the PCR clone pSB17 (section 3.3.3), probes 2 and three were *Xba* I -*Sca* I and *Xba* I-*Hind* III restriction fragments obtained from an inverse PCR clone (BINV3) of the proteolytic type B neurotoxin (Dr S.M. Whelan., Ph.D thesis 1993). Probe 4 was the entire BINV3 PCR clone.

3.3.3 Creation of a genomic restriction map of the silent neurotoxin gene N-terminal coding region.

Nucleotide sequence data derived from pSB17 illuminated the requirement for a genomic restriction map to encompass the silent neurotoxin gene and a considerable 5' region. To ensure that an 'upstream' clone had originated from the silent B locus and not a type A locus, it would have to be selected for and be contiguous with a type B sequence.

A 395 bp *Bgl*II-*Stu*I restriction fragment was isolated from pSB17. This fragment was essentially homologous with the proteolytic type B neurotoxin upstream sequence and did not contain tracts of DNA identical to the type A upstream sequence (figure 3.2). The choice of this restriction fragment in combination with *Bgl*II double digests and *Hae*III double digests (*Hae*III is within the *Stu*I palindrome) allowed the precise delineation of restriction sites 5' and 3' to this probe's region (figure 3.2). Samples of *C. botulinum* NCTC 2916 genomic DNA were digested with a number of restriction enzymes and the region 5' to the putative serotype B gene was mapped using the *Bgl*II-*Stu*I probe by Southern blotting (figure 4.7a and 4.7b). A 6.5 kb *Xba*I fragment identified in the map was targeted for cloning the N-terminal coding region of the silent neurotoxin gene and at least 4 kb of upstream DNA (figure 3.2).

3.3.4 Creation of a genomic restriction map of the silent neurotoxin gene.

Due to the large size of a clostridial neurotoxin gene, a more extensive area was mapped, encompassing all of the neurotoxin coding region. Probing of double digests of *C. botulinum* NCTC 2916 genomic DNA, generated with combinations of either *Hae*III or *Hind*III enzymes using a *Sca*I- *Xba*I DNA restriction fragment derived from a proteolytic type B neurotoxin clone BINV3 (Whelan *et al.*, 1992), extended and improved the existing restriction map (figure 3.2). However a comparison of the restriction profile with the published proteolytic type B neurotoxin sequence highlighted some variation, specifically the *Hind*III maps did not correlate. A restriction enzyme site which did appear to occur at a convenient spacing was *Xba*I. Verification of the *Xba*I sites in relation to the other mapped restriction sites was achieved by probing a variety of *Xba*I double digests with a *Hind*III-*Xba*I probe, which was also obtained from the BINV3 clone (Whelan *et al.*, 1992). Thus a further two contiguous *Xba*I

fragments of 1.6 and 1.9 kb were targeted for cloning which were potentially capable of encoding the entire silent neurotoxin gene and adjacent areas.

3.3.5 Cloning the silent neurotoxin gene.

Approximately 500 µg of *C. botulinum* NCTC 2916 genomic DNA was digested with 400 units of *Xba*I for 5 h. The sample was separated in a 0.8 % (w/v) agarose gel before sections of the gel were excised which corresponded approximately to 6.5, 1.9 and 1.6 kb. The DNA was recovered from the gel by electroelution, purified by phenol extraction and ligated into the *E. coli* cloning vector pMTL22 (Chambers *et al.*, 1988). In total, 960 transformant colonies from each respective fragment population were probed for the presence of an insert which hybridised with either the *Bgl*II-*Stu*I restriction fragment or a *Hind*III restriction fragment which bridged the 3' *Xba*I site in the proteolytic type B neurotoxin (figure 3.2). Representative clones for each potential fragment were isolated and restriction analyses of purified plasmid DNA with *Xba*I revealed the presence of inserts of the predicted size. It was interesting to note that clones containing the two larger inserts were blue in the presence of X-gal and IPTG. Clones were designated pSB1, pSB2 and pSB3 according to their relative position (figure 3.2).

3.3.6 Nucleotide sequence analysis of silent neurotoxin gene in *C. botulinum* NCTC 2916.

Preliminary sequence data obtained for pSB1 using a reverse universal primer confirmed the presence of a type B-like neurotoxin sequence, whose reading frame read in an opposite direction to the *lacZ* α gene into which it was cloned. The sequence data using a forward universal primer suggested the presence of another component of the neurotoxin complex (section 4.2.1.6). The entire insert was subcloned as an *Xba*I restriction fragment into the M13mp18 bacteriophage vector to facilitate nucleotide sequencing. Sequence data from the distal ends of the pSB2 insert also suggested the presence of a type B-like neurotoxin gene. Both *Xba*I sites appeared to correspond to those present in the published proteolytic type B neurotoxin sequence suggesting that the inserts of pSB1 and pSB2 were contiguous. In a similar vein, the sequence obtained from pSB3 with forward universal primer appeared to correlate with the sequence

adjacent to the 3' *Xba*I site in the published proteolytic type B neurotoxin sequence. In conclusion, the three contiguous clones were predicted to contain the equivalent coding region of the entire silent neurotoxin gene.

pSB3 was cloned into the M13mp18 bacteriophage vector as a *Xba*I fragment. All three clones were sequenced on both strands using the manual dideoxy chain termination method with custom made oligonucleotide primers. Primers were designed from the published proteolytic type B neurotoxin sequence (Whelan *et al.*, 1992). However when a deviation in the silent neurotoxin sequence affected primer annealing, primers were designed from the raw sequence data. Data was analysed and assembled using DNASTAR giving a single contig. The greatest problem in generating over 4 kb of sequence data by the 'manual' dideoxy-Sanger method was in determining a reliable method to prevent the acrylamide gel from adhering to both glass plates during electrophoresis.

3.3.7 Comparison of the silent neurotoxin gene in *C. botulinum* NCTC 2916 with the published proteolytic and non-proteolytic type B neurotoxin sequences.

3.3.7.1 Nucleotide sequence.

An initial comparison of the complete nucleotide sequence obtained from clones pSB1, pSB2 and pSB3 with the published proteolytic type B neurotoxin gene (Whelan *et al.*, 1992) suggested that a complete neurotoxin gene was present (figure 3.3). The A+T DNA content for the 'coding' region was 74.78 % which was marginally higher than that for both of the other type B genes which shared an A+T DNA content of 74.51 %. Alignment of the sequence with both type B genes using the Martinez/Needleman-Wunch method (MEGALIGN programme, DNASTAR Inc.) demonstrated identity of 97.5 % and 94.8 % with the proteolytic and nonproteolytic genes respectively. This indicated that over approximately 3.9 kb, significant, but not dramatic, variations occurred between the genes and that overall, the silent gene appeared to be most similar to the proteolytic type B neurotoxin sequence.

A close inspection of the silent gene showed that 6 contiguous nucleotides were missing, starting at nucleotide position 983 (figure 3.3). This deletion removed 2 codons encoding lysine and tyrosine, and is not present in either of the type B neurotoxin genes.

BOTB^P TGC GCATTTATGGGCATTAAAAGGGATATAAACTTAAAATAAGGAGGAGAATATTT CCAGTTACAATAAATAATTTT 80
BOTB^O ---A-----
BOTB^N ---A-----A-----

BOTB^P AATTATAATGATCCTATTGATAATAATAATATTATTATGATGGAGCCTCCATTTGCGAGAGGTACGGGGAGATATTATAA 160
BOTB^O -----A-----T-----
BOTB^N -----G-C-----A-----A-G-----

BOTB^P AGCTTTTAAAATCACAGATCGTATTTGGATAATACCGGAAAGATATACTTTTGGATATAAACCTGAGGATTTTAATAAAA 240
BOTB^O -----T-----
BOTB^N -----C-----

BOTB^P GTTCCGGTATTTTAAATAGAGATGTTTGTGAATATTATGATCCAGATTACTTAAATACTAATGATAAAAAGAATATATTT 320
BOTB^O -----G-----
BOTB^N -----C-----

BOTB^P TTACAAACAATGATCAAGTTATTTAATAGAATCAAATCAAACCATTTGGGTGAAAAGTTATTAGAGATGATTATAAATGG 400
BOTB^O -----
BOTB^N --C-----T-----

BOTB^P TATACCTTATCTTGGAGATAGACGTGTTCCACTCGAAGAGTTTAACACAAACATTGCTAGTGTAAGTGTTAATAAATTAA 480
BOTB^O -----T-----
BOTB^N -----

BOTB^P TCAGTAATCCAGGAGAAGTGGAGCGAAAAAAGGTATTTTCGCAAATTTAATAATATTTGGACCTGGGCCAGTTTTAAAT 560
BOTB^O -----G-----
BOTB^N -T-----

BOTB^P GAAATGAGACTATAGATATAGGTATACAAAATCATTTTGCATCAAGGGAAGGCTTCGGGGGTATAATGCAAATGAAGTT 640
BOTB^O -----C-----
BOTB^N -----T-----A--

BOTB^P TTGCCCAGAATATGTAAGCGTATTTAATAATGTTCAAGAAAACAAAGGCGCAAGTATATTTAATAGACGTGGATATTTTT 720
BOTB^O -----
BOTB^N ---T-----

BOTB^P CAGATCCAGCCTTGATATTAATGCATGAACTTATACATGTTTTACATGGATTATATGGCATTAAAGTAGATGATTTACCA 800
BOTB^O -----C-----T----A-----
BOTB^N -----G-----

BOTB^P ATTGTACCAAATGAAAAAAATTTTTTATGCAATCTACAGATGCTATACAGGCAGAAGAACTATATACATTTGGAGGACA 880
BOTB^O ----G-----G--
BOTB^N -----A-----

BOTB^P AGATCCCAGCATCATAACTCCTTCTACGGATAAAAGTATCTATGATAAAGTTTTCGAAAATTTTAGAGGGATAGTTGATA 960
BOTB^O -----G-----
BOTB^N -----T-----A-----G-----

BOTB^P GACTTAACAAGGTTTTAGTTTGCATATCAGATCCTAACATTAATATTAATATATATAAAAATAAATTTAAAGATAAATAT 1040
BOTB^O -----ΔΔ-----
BOTB^N -----C-----

BOTB^P AAATTCGTTGAAGATTCTGAGGGAAAATATAGTATAGATGTAGAAAGTTTTGATAAATTATATAAAAGCTTAATGTTTGG 1120
BOTB^O ΔΔΔΔ-----A-----C-----
BOTB^N -----A-----C-----A--

BOTB^P TTTTACAGAACTAATATAGCAGAAAATTATAAAATAAAAACTAGAGCTTCTTATTTTAGTGATTCCTTACCACCAGTAA 1200
BOTB^O -----
BOTB^N -----T-----

BOTB^P AAATAAAAAATTTATTAGATAATGAAATCTATACTATAGAGGAAGGGTTTAATATATCTGATAAAGATATGGAAAAAGAA 1280
BOTB^O -----A-----
BOTB^N -----A-----G-----

BOTB^P TATAGAGGTCAGAATAAAGCTATAAATAAACAAGCTTATGAAGAAATTAGCAAGGAGCATTGCTGTATATAAGATACA 1360
BOTB^O -----C-----A-----
BOTB^N -----G-----C-----

BOTB^P AATGTGTAAAAGTGTTAAAGCTCCAGGAATATGTATTGATGTTGATAATGAAGATTTGTTCTTTATAGCTGATAAAAAATA 1440
BOTB^O -----G-----
BOTB^N -----T-----C-----A-----

BOTB^P GTTTTTCAGATGATTTATCTAAAAACGAAAGAATAGAATATAATACACAGAGTAATTATATAGAAAATGACTTCCCTATA1520
BOTB^o -----C-----A-----T---T-----
BOTB^N -----T-----G-----A-----G-----T-----

BOTB^P AATGAATTAATTTTAGATACTGATTTAATAAGTAAAATAGAATTACCAAGTGAAAATACAGAATCACTTACTGATTTTAA1600
BOTB^o -----
BOTB^N -----

BOTB^P TGTAGATGTTCCAGTATATGAAAAACAACCCGCTATAAAAAAATTTTACAGATGAAAATACCATCTTCAATATTTAT1680
BOTB^o ---T-----A---A-----T-----
BOTB^N -----G-----

BOTB^P ACTCTCAGACATTTCTCTAGATATAAGAGATATAAGTTTAACATCTTCATTTGATGATGCATTATTATTTTCTAACAAA1760
BOTB^o -T-----
BOTB^N -----A-----G-----G-----

BOTB^P GTTTATTCATTTTTTTCTATGGATTATATTAAACTGCTAATAAAGTGGTAGAAGCAGGATTATTTGCAGGTTGGGTGAA1840
BOTB^o -----
BOTB^N -----A-----

BOTB^P ACAGATAGTAAATGATTTTGTAATCGAAGCTAATAAAAGCAATACTATGGATAAAATTGCAGATATATCTCTAATTGTTTC1920
BOTB^o -----G-----G-----
BOTB^N -----G-----G-----

BOTB^P CTTATATAGGATTAGCTTTAAATGTAGGAAATGAAACAGCTAAAGGAAATTTTGAAAATGCTTTTGAGATTGCAGGAGCC2000
BOTB^o -----
BOTB^N -----G-----G-----T--

BOTB^P AGTATTCTACTAGAATTTATACCAGAACTTTTAATACCTGTAGTTGGAGCCTTTTTATTAGAATCATATATTGACAATAA2080
BOTB^o -----T-----
BOTB^N -----T-----

BOTB^P AAATAAAATTATTAAACAATAGATAATGCTTTAACTAAAAGAAATGAAAATGGAGTGATATGTACGGATTAATAGTAG2160
BOTB^o -----G-----A---G-----G-----T-----
BOTB^N -----GTG-----T-----

BOTB^P CGCAATGGCTCTCAACAGTTAATACTCAATTTTATACAATAAAAGAGGGAATGTATAAGGCTTTAAATTATCAAGCACAA2240
BOTB^o -----
BOTB^N -----

BOTB^P GCATTGGAAGAAATAATAAAATACAGATATAATATATATTCTGAAAAAGAAAAGTCAAATATTAACATCGATTTTAATGA2320
BOTB^o -----A-----G-----G-----
BOTB^N -----A-----G-G-----A-----

BOTB^P TATAAATTCTAAACTTAATGAGGGTATTAACCAAGCTATAGATAATATAAATAATTTTATAAATGGATGTTCTGTATCAT2400
BOTB^o -G-----A-----
BOTB^N -----T-----G-----G-----A-----

BOTB^P ATTTAATGAAAAAATGATTCCATTAGCTGTAGAAAAATTACTAGACTTTGATAATACTCTCAAAAAAATTTGTTAAAT2580
BOTB^o -----C-----G-----A-----
BOTB^N -----A-----A-----

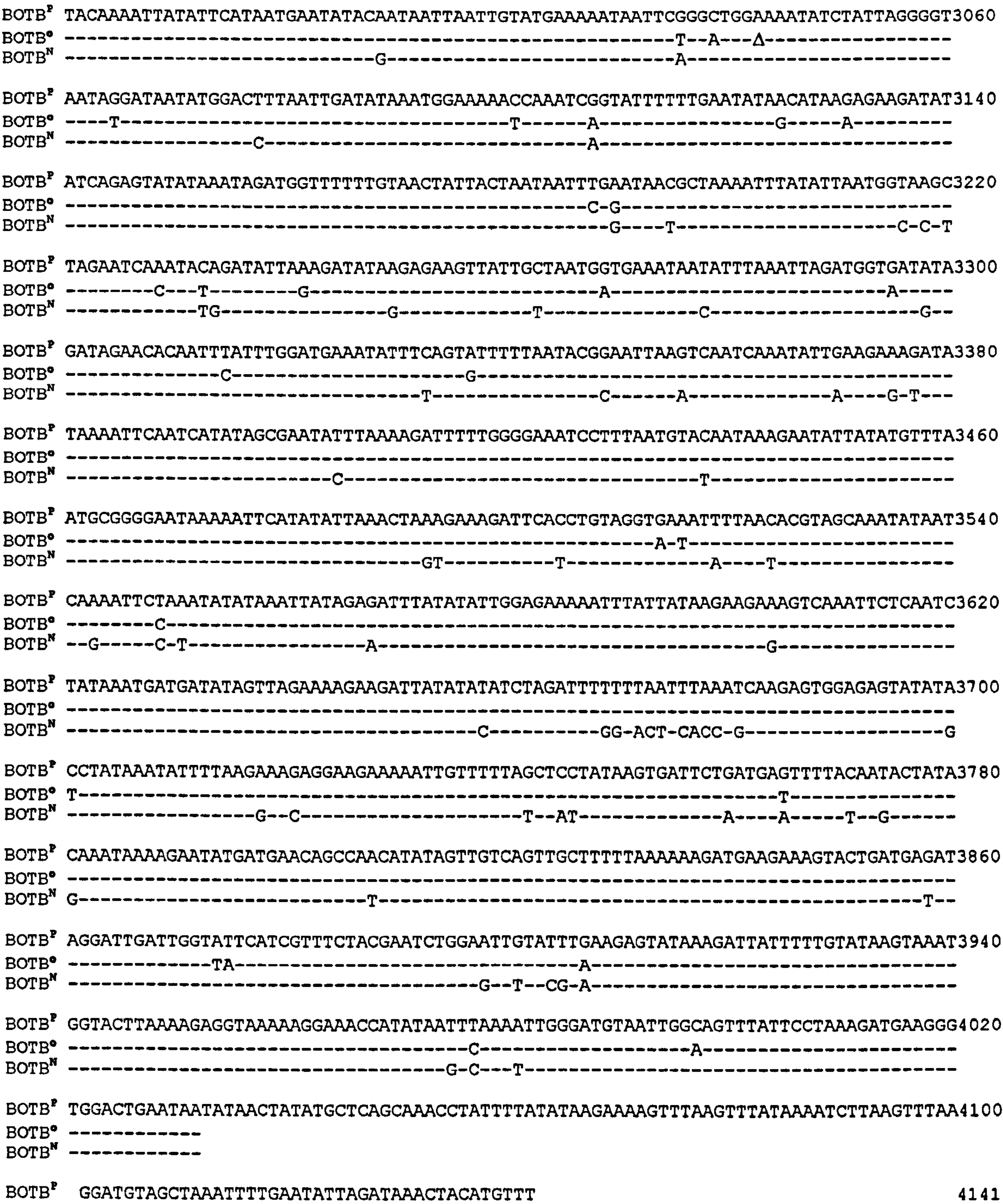
BOTB^P TATATAGATGAAAATAAATTATATTTGATTGGAAGTGCAGAATATGAAAATCAAAGTAAATAAATACTTGAAAACCAT2660
BOTB^o -----G-----G-----
BOTB^N -----A-----T---G-----G-----

BOTB^P TATGCCGTTTGATCTTTCAATATATACCAATGATACAATACTAATAGAAATGTTTAATAAATATAATAGCGAAATTTTAA2740
BOTB^o ---A-----G---T---A-----A-----T-----
BOTB^N ---A-A-----CG---T-T---AT-GA-----A---A-----

BOTB^P ATAATATTATCTTAAATTTAAGATATAAGGATAATAATTTAATAGATTTATCAGGATATGGGGCAAAGGTAGAGGTATAT2820
BOTB^o -----G-----G-----
BOTB^N -----GA-----A-----

BOTB^P GATGGAGTCGAGCTTAATGATAAAAAATCAATTTAAATTAAGTTCAGCAAATAGTAAGATTAGAGTGACTCAAATCA2900
BOTB^o -----G---A-----T-----
BOTB^N -----G---A-----G-----C-----

BOTB^P GAATATCATATTTAATAGTGTGTTTCCTTGATTTTAGCGTTAGCTTTTGGATAAGAATACCTAAATATAAGAATGATGGTA2980
BOTB^o -----A-----T-----
BOTB^N -----T-----A-----G-----G-----A---



4141

Figure 3.3 Comparative alignment of the silent BoNT/B gene.

The silent type B neurotoxin gene BOTB[°], is aligned with the published proteolytic type B neurotoxin gene BOTB^P (Whelan *et al.*, 1992) and the non-proteolytic type B neurotoxin gene BOTB^N (Hutson *et al.*, 1994). Conserved nucleotides are marked with (-) whilst deleted residues are marked with (Δ). Nucleotide numbering is relative to BoNT/B^P. The thymine residue which creates a nonsense stop codon is shown in magenta, ribosome binding site in red and initiation codon in yellow.

A second single base pair deletion occurs at position 2881. Here there are only 3 adenine residues instead of a run of four. Of the remaining 86 base substitutions, 26 produce conservative mutations, 24 produce semi-conservative amino acid substitutions, 37 produce non-conservative amino acid changes using the groupings (ILV, ST, EDNQ, WYF, KR, P, A, C, H, M and G) and one causes the formation of an amber stop codon (UAG) from a glutamic acid codon (GAG) at position 383 or codon 128 (figure 3.4). Among these base changes, the stop codon at position 128 of the silent neurotoxin gene would, in the absence of any other changes, entirely account for the observed 'silent' type B neurotoxin phenotype. Similarly, the frame-shift alone would produce a truncated protein with a nonsense C-terminal sequence. However, characterisation of all the amino acid changes may also reveal more subtle influences on the structure of this silent gene. If the stop codon and frame shift in the silent neurotoxin gene were 'corrected', the light chain would contain 8 conservative codon changes, 4 semi-conservative amino acid changes and 5 non-conservative amino acid changes, using the amino acid groupings above, whereas the heavy chain would contain 18 conservative changes, 18 semi-conservative changes and 32 non-conservative changes. The deviations in both nucleic acid and amino acid sequence appear to have a disproportional bias towards the heavy chain, whilst changes in the light chain are proportionally under represented by 50 %.

Determination of the position of the nucleotide substitutions within each codon between the silent gene and proteolytic type B neurotoxin gene demonstrates that overall, no particular position has a greater frequency of substitution, indeed the distribution appears to be particularly constant (table 3.1). To test whether this seemingly random substitution was associated with the silent gene or if it were a common feature of other closely related genes, the exact position of nucleotide substitutions was determined in two other closely related pairs of clostridial neurotoxins, for which a complete nucleotide sequence has been determined (table 3.1).

PROB MPVTINNFNYNIDPIDNNNIIMMEPPFARGTGRIYKAFKITDRIWII PERYTFGYKPEDFNKSSGIFNRDV	70
SILB -----M-----L-----	
NPRB -----D-----	
PROB CEYYDPDYLNTNDKKNIFLQTMIKLENRIKSKPLGEKLLEMIINGIPYLGDRRVPLEE FNTNIASVTVNK	140
SILB W-----*	
NPRB -----F--L-----	
PROB LISNPGEVERKKGIFANLIIFGPGPVLNENETIDIGIQNHFA SREGFGGIMQMKECPEYVSVFNNVQENK	210
SILB -----A-----	
NPRB -----	
PROB GASIFNRRGYFSDPALILMHელიHVLHGLYGIKVDDLPIVPNEKKFFMQSTDAIQAEELYTFGGQDPSII	280
SILB -----I-N-----	
NPRB -----T-----	
PROB TPSTDKSIYDKVLQNFGRGIVDRLNKVLVCISDPNININIYKNKFKDKYKFVEDSEGKYSIDVESFDKLYK	350
SILB S-----ΔΔ-----	
NPRB S-----N-----	
PROB SLMFGFTETNIAENYKIKTRASYFSDSLPPVKIKNLLDNEIYTTIEEGFNISDKDMEKEYRGQNKAINKQA	420
SILB -----N-----	
NPRB ---L---I-----N-G-----	
PROB YEEISKEHLAVYKIQMCKSVKAPGICIDVDNEDLFFIADKNSFSDDL SKNERIEYNTQSNYIENDFPINE	490
SILB -----I-----A---N---S---	
NPRB -----V-----N-----V---N---G-----	
PROB LILDTDLISKIELPSENTESLTDENVDPVYEQPAIKKI FT DENTIFQYLYSQT FPLDIRDISLTSSFD	560
SILB -----Y--E-K-L-----	
NPRB -----V-----N-----	
PROB DALLFSNKVYSFFSMDYIKTANKVVEAGLFAGWVKQIVNDFVIEANKSNTMDKIADISLIVPYIGLALNV	630
SILB -----D-----S-----	
NPRB ----V-S-----D-----S-----	
PROB GNETAKGNFENAFEIAGASILLEFIPELLIPVVGAFLL ESYIDNKNKIIKTIDNALT KRNEKWSDMYGLI	700
SILB -----E--NS---D---I-----	
NPRB -D-----S-----S-----V-----V---I-----	
PROB VAQWLSTVNTQFYTIKEGMYKALNYQAQALEEIIKYRYNIYSEKEKSNINIDFNDINSKLN EGINQAIDN	770
SILB -----K-----R-----V-----	
NPRB -----K-----E-----N-----D-----M---	
PROB INNFINGCSVSYLMKKMIPLAVEKLLDFDNTLKKNLLNYIDENKLYLIGSAEYEKSKVNKYLKTIMPFDL	840
SILB -----E-----R-----D-----SI-----	
NPRB --D---E-----K-----V-D---D-----I-----	
PROB SIYTNDTILIE MENKYNSEILNNIILNRYKDNNLIDLSGYGAKVEVDGVELNDKNQFKLTSSANSKIR	910
SILB -M---N---I---D---R--K-----K-----	
NPRB -T-S-IE---KI-----R-----K-----D-----	
PROB VTQONQNIIFNSVFLDFSVSFWIRIPKYKNDGIQNYIHNEYTIINCMKNNSGWKISIRGNRIIWT LIDING	980
SILB -I-----M-----Δ---M-----	
NPRB -----M-----R--D-----	
PROB KTKSVFFEYNIREDISEYINRWFFVTITNNLNNAKIYINGKLESNTDIKDIREVIANGEIIFKLDGDIDR	1050
SILB -I-----S-K-----SD-----HI--R-----D-----N---	
NPRB -----D-----T---M---G---V---T-----V---	
PROB TQFIWMKYFSIFNTELSQSNIERYKIQSYSEYKDFWGNPLMYNKEYYMFNAGNKNSYIKLKKDSPVGE	1120
SILB -----V-----	
NPRB -----Q-N---K-I-----V---S-----	
PROB ILTRSKYNQNSKYINYRDLYIGEKFII RRKSNSQSINDDIVRKEDYIYLDFFNLNQEWRVYTYKYFKKEE	1190
SILB -----I-----	
NPRB --I-----N---N-----E-----H-LVLHHE-----A-----EQ-	
PROB EKLFLAPISDSDEFYNTIQIKEYDEQPTYSCQLLFKKDEESTDEIGLIGIHRFYESGIVFEEYKDYFCIS	1260
SILB -----D-----Y-----K-----	
NPRB -----SI---N---K--E-----S-----D-----VLRKK-----	
PROB KWYLKEVKRKPYNLKLGCNWQFIPKDEGWTE .	1291
SILB -----S-----K----- .	
NPRB -----KSN----- .	

Figure 3.4 Comparative alignment of the type B neurotoxin amino acid sequences.

The amino acid sequences for the proteolytic type B neurotoxin (PROB), (Whelan *et al.*, 1992) and non-proteolytic type B neurotoxin (NPRB), (Hutson *et al.*, 1994) are compared. Amino acids are indicated by the single letter code. Dashes (-) represent amino acids with identity. The silent neurotoxin gene (SILB) is altered by replacing the nonsense stop codon with a *, whole codon deletions with a Δ and single nucleotide deletion which would otherwise disrupt the amino acid reading frame with a Δ . The amino acid positions indicated on the right correspond to the proteolytic type B sequence. Features of particular interest including the proposed active site region; HELIHVLHGLYG (aa 230-241), light chain-heavy chain junction; VK (aa 440-441) and cysteine residues implicated in the light chain-heavy chain disulphide bond (aa 437 and 446) are shown in cyan.

NEUROTOXIN GENE PAIR	CODON POSITION I	CODON POSITION II	CODON POSITION III
Silent B vs. Proteolytic B	30.2 %	33.8 %	36.0 %
Non-proteolytic B vs. Proteolytic B	36.8 %	20.6 %	42.6 %
<i>C. botulinum</i> E vs. <i>C. butyricum</i> E	39.5 %	44.7 %	15.8 %

Table 3.1 Distribution of nucleotide substitutions within the codons of three pairs of neurotoxin genes.

The neurotoxin genes are the silent neurotoxin gene, (this study), the proteolytic type B neurotoxin, (Whelan *et al.*, 1992), the non-proteolytic type B neurotoxin, (Hutson *et al.*, 1994) and the type E neurotoxins from *C. botulinum* and *C. butyricum* (Poulet *et al.*, 1992). The comparisons show that, for this limited number of genes, the distribution of nucleotide substitutions within codons alters between the toxin gene serotypes examined.

3.3.7.2 Amino acid sequence.

Alignment of the stop codon-corrected amino acid sequence of the silent B neurotoxin with examples of all known serotypes of botulinal neurotoxins and tetanus neurotoxin (Elmore *et al.*, 1995) allowed a rational speculation as to the nature of the amino acid changes found between the silent neurotoxin primary sequence, the type B sequences and the other serotypes. Amino acid changes appeared to fall into four groups. Firstly, there were two changes which occurred in regions which were absolutely conserved in all other clostridial neurotoxins. Both of these occurred in the light chain region. The non-conservative replacement of a proline for a leucine residue at position 47 occurs in a highly conserved motif containing two other conserved residues. A second change where contiguous lysine and tyrosine residues are deleted occurs at positions 327-8, removing this short motif which is also found in all other clostridial neurotoxins (figure 3.5).

Y K N K F K D - - K F V E D S E G (329)	BoNT/B (SILENT)
M K N V F K E K Y L L S E D T S G 329	BoNT/A
M K N V F K E K Y L L S E D T S G 329	BoNT/A (Inf)
Y K N K F K D K Y K F V E D S E G 336	BoNT/B (GpI)
Y K N K F K D K Y K F V E D S E G 336	BoNT/B (GpII)
Y K Q K L I R K Y R F V V E S S G 338	BoNT/C
Y K K Y F S E K Y N F D K D N T G 338	BoNT/D
Y K D V F E A K Y G L D K D A S G 315	BoNT/E
Y K D V F E A K Y G L D K D A S G 315	BoNT/E (Buty)
Y K D Y F Q W K Y G L D K N A D G 332	BoNT/F (GpI)
Y K D Y F Q W K Y G L D K N A D G 332	BoNT/F (GpII)
Y K N F F Q W K Y G L D Q D S N G 324	BoNT/F (Barat)
Y K Q I Y K N K Y D F V E D P N G 335	BoNT/G
Y K Q I Y Q Q K Y Q F D K D S N G 338	TeNT

Figure 3.5 Comparative alignment of a clostridial neurotoxin motif absent in BoNT/B (silent)

Alignment of all known clostridial neurotoxins demonstrates that a number of amino acids are totally conserved (Elmore *et al.*, 1995). The comparison above shows that a Lysine-Tyrosine two amino acid motif is found in all of the neurotoxin primary amino acid sequences, but is deleted in the silent type B neurotoxin’s ‘coding sequence’. This type of aberation may render any pharmacological potential of the silent type B neurotoxin redundant.

A second class of amino acid substitutions can be grouped on the basis that these differences only occur between the proteolytic type B neurotoxin and silent neurotoxin, but there is no equivalent difference between the non-proteolytic type B neurotoxin and silent neurotoxin. These substitutions would logically be formed by changes to the proteolytic type B sequence after the evolutionary separation of these two genes. Fourteen such changes exist and may reflect the duration that these sequences have been independent.

The third group is composed of amino acid changes which are different from both proteolytic and non-proteolytic type B neurotoxins, but are widely used at that position by most of the other toxin serotypes (Elmore *et al.*, 1995). Although only 3 of these are semi-conservative, the other 8 could be considered as semi-conservative as they may fulfill a similar biological role in the other toxin serotypes. A fourth group exists which falls into neither of the 3 previously described categories and seems to be composed of residue changes, which based on neurotoxin sequence similarities (Elmore *et al.*, 1995), occur in between conserved motifs or rather in the 'variable' regions of the neurotoxin sequences. This group contains both semi-conserved and non-conserved residues.

3.3.8 Comparison of the 5 prime region

Analysis of the nucleotide sequence 5' of the predicted initiation codon suggested that the silent gene was preceded by another coding region, separated by a 25 bp intergenic region similar to the proteolytic and non-proteolytic type B neurotoxin genes (Whelan *et al.*, 1992; Hutson *et al.*, 1994). The intergenic region in front of the silent neurotoxin gene was identical to both type B sequences and contained the polypurine sequence previously proposed for the ribosome binding site (Whelan *et al.*, 1992).

3.4 Transcriptional mapping of the silent B gene using Northern blotting and primer transcript analysis

Total RNA was isolated from *C. botulinum* NCTC 2916, *C. botulinum* NCTC 7272 (a type A positive control) and *C. botulinum* type B strain Danish, the proteolytic neurotoxin which has its nucleotide sequence determined (Whelan *et al.*, 1992).

3.4.1 Northern analysis

RNA was analysed by agarose gel electrophoreses and blotted onto a Nylon membrane and probed with a type A neurotoxin and type B neurotoxin specific DNA sequence (figure 3.6a and 3.6b). The blots clearly demonstrate that *C. botulinum* NCTC 2916 contains both a type A and a type B neurotoxin transcript, whilst the *C. botulinum* NCTC 7272 only contains a type A transcript and *C. botulinum* Danish only contains a type B neurotoxin transcript.

Inspection of the blots (figure 3.6a and 3.6b) suggested that the strongest signal originated from a point immediately below the 23S rRNA band at a size of about 2.6 kb. However, the ribosomal RNA bands appeared to interfere with the binding of the probe to the mRNA transcript, presumably at high concentration, possibly by competing for ligands when binding to the Nylon membrane. In light of this supposition, the transcript could be deemed to be somewhat larger. Indeed, most of the neurotoxin transcripts in these blots appear to form a smear which initiates from a quite faint but distinct band of at least 7.4 kb, most clearly seen on figure 3.6b. This appears to be the case for type A, type B and the 'silent' neurotoxin transcript. A similar artifact was previously noticed in Northern blots for the type A neurotoxin of *C. botulinum* type A (Dr I. Henderson, pers. comm.) and can be seen in blots of *C. difficile* RNA (Hammond *et al.*, 1997).

3.4.2 Primer transcript analysis.

Attempts were made to map a transcription initiation site for the silent type B gene by primer extension using primer 3237 (Appendix 5) from both *C. botulinum* NCTC 2916 total RNA and *C. botulinum* type B Danish total RNA. A number of bands were produced. However, the majority of the bands did not appear in the same position in both strains (figure 3.7), some bands appeared in the neurotoxin coding region whilst only 2 bands coincided in both strains. These two bands occurred close to the start of the neurotoxin genes and the strongest were positioned 5 bp in front of the ribosome binding site of the neurotoxin gene. This result did not give a satisfactory basis from which a conclusion could be drawn. It was not clear if the multiple bands represented separate initiation points of an mRNA species or if a degree of false priming with other

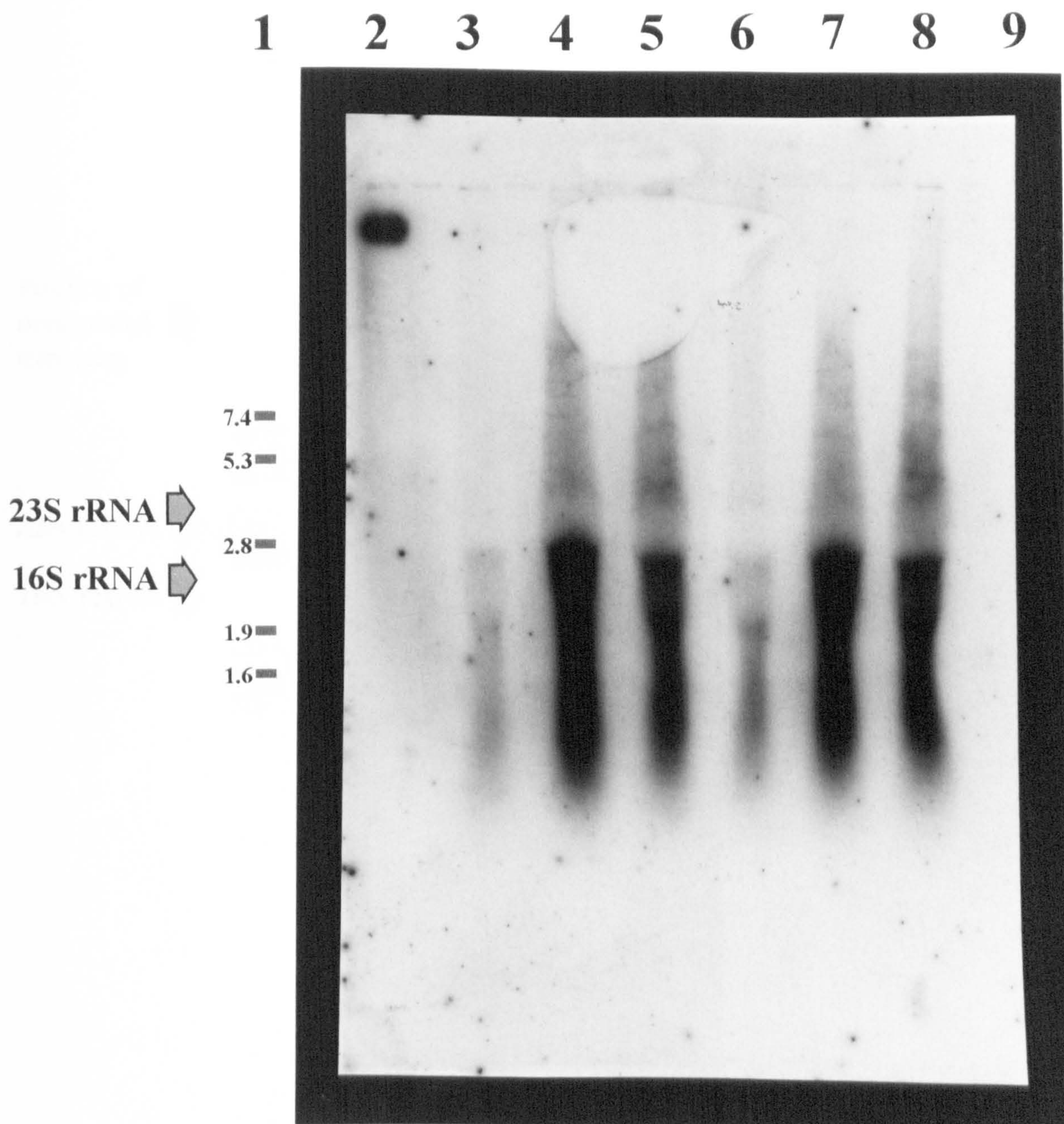


Figure 3.6a Northern blot of *C. botulinum* total RNA probed with a type A neurotoxin specific probe

Total RNA isolated from *C. botulinum* strains was analysed by Northern blotting (section 3.4). Lane 1) RNA molecular weight markers (Boehringer Mannheim) 2) type A DNA positive control 3) Strain B Danish 4) NCTC 2916 (A/B^o) 5) NCTC 7272 (A) 6) Strain B Danish 7) NCTC 2916 (A/B^o) 8) NCTC 7272 (A) 9) type B positive DNA control. Probing the fixed RNA samples with a type A neurotoxin specific radiolabeled DNA fragment generated using type A neurotoxin serotype specific PCR primers (Whelan, S. M. 1993) detected a type A transcript in *C. botulinum* NCTC 7272, a strain only known to contain a type A neurotoxin gene, and *C. botulinum* NCTC 2916 a type A neurotoxin strain harbouring a type B neurotoxin gene. RNA from the type B strain did not hybridise significantly with the type A neurotoxin probe.

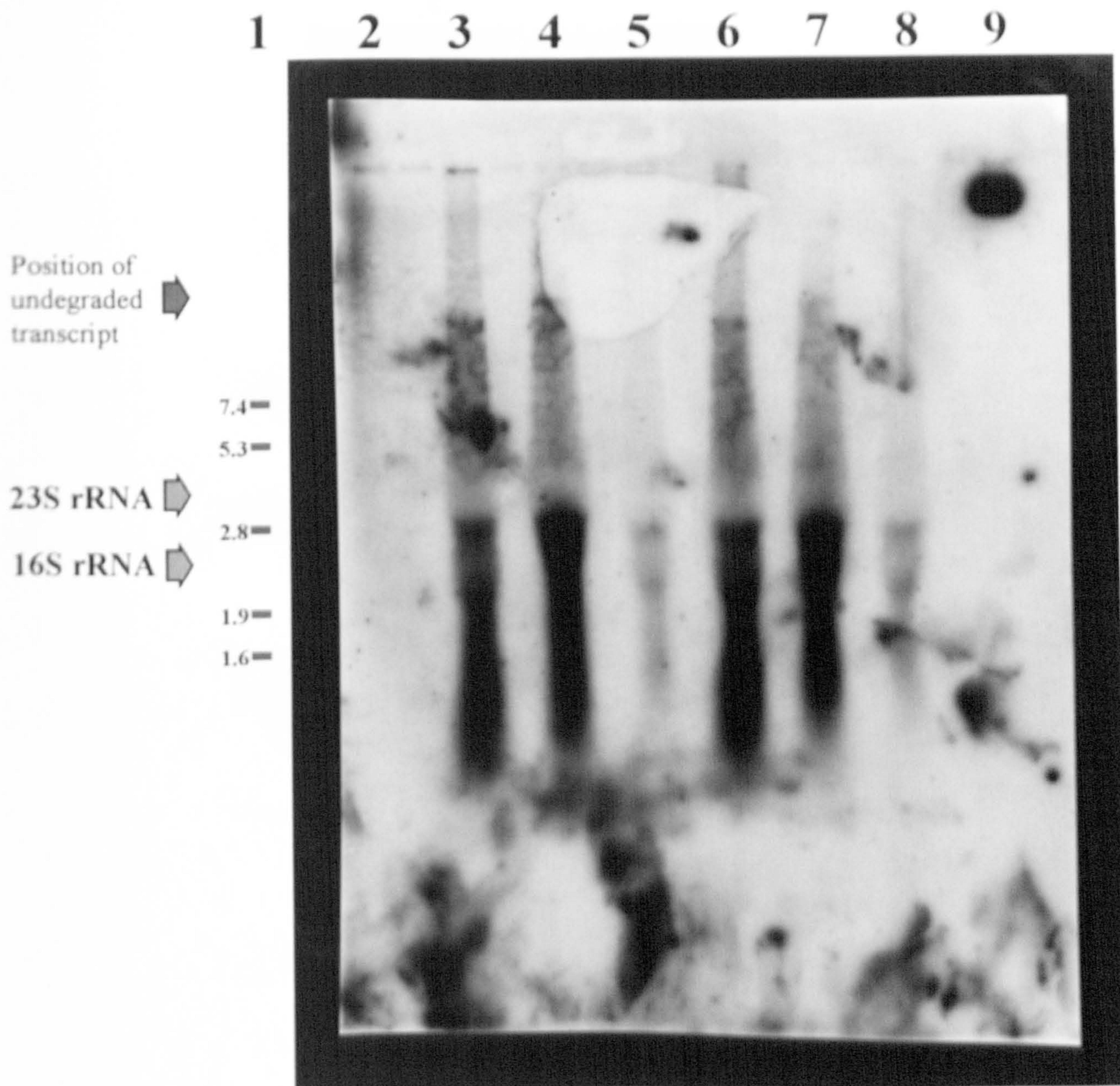


Figure 3.6b Northern blot of *C. botulinum* total RNA probed with a type B neurotoxin specific probe

Total RNA isolated from *C. botulinum* strains was analysed by Northern blotting (section 3.4). Lane 1) RNA molecular weight markers (Boehringer Mannheim) 2) type A DNA positive control 3) Strain B Danish 4) NCTC 2916(A/B^o) 5) NCTC 7272 (A) 6) Strain B Danish 7) NCTC 2916 (A/B^o) 8) NCTC 7272 (A) 9) type B positive DNA control.

Probing the fixed RNA samples with a type B neurotoxin specific radiolabeled DNA fragment generated using type B neurotoxin serotype specific PCR primers (Whelan, S. M. 1993) detected a type B transcript in *C. botulinum* NCTC 7272, a strain only known to contain a type B neurotoxin gene, and *C. botulinum* NCTC 2916, a type A neurotoxin strain harbouring a type B neurotoxin gene. RNA from the type A strain *C. botulinum* NCTC 7272 did not hybridise significantly with the type B neurotoxin probe.

A G C T 1 2

5' A T A T A A A C T T A A A A T A A G G A G G A G A A T A T T T A T G 3'



Figure 3.7 Detection of 5' mRNA transcripts by primer extension

Figure 3.7 Detection of 5'mRNA transcript ends by primer extension

Transcripts were mapped by generating cDNA from radiolabeled oligonucleotide primers using the method detailed in section 3.4.2. Total RNA samples from *C. botulinum* NCTC 2916 and the proteolytic type B strain (Danish) were both extended with primer 3273 (Appendix V). Primer extension products were separated by denaturing gradient polyacrylamide gel electrophoresis beside the respective plasmid sequencing reaction (lanes AGCT). Lane 1 was cDNA products generated from NCTC 2916 whilst lane 2 was the cDNA product generated from strain Danish. The accompanying sequence is of the intergenic gap between the NTN_H gene and the silent BoNT/B where the stop codon of the NTN_H gene is blue, the base corresponding to the primer extension products in both standard and primer extension-blotting is tan, the ribosome binding site in red and the initiation codon of the silent type B neurotoxin gene in green.

transcripts had occurred. To produce an answer to these questions an alternative strategy was devised.

3.4.3 A novel modification for primer transcript analysis.

To avoid the possibility of generating radio-labeled primer transcripts from mRNA other than the type B neurotoxins, the primer extension reactions (section 2.2.10.3) were conducted with unlabeled oligonucleotides 3237 and 2677 (Appendix 5). The cDNA products were then separated on a polyacrylamide gel (section 2.2.5.4) next to a non-radiolabeled dideoxy-sequencing reaction using primers 3237, 2677 and a plasmid clone encoding the 5' region of the neurotoxin gene. The nucleic acids were then transferred to a positively charged Nylon membrane via Southern blotting. Alkali fixing of the cDNA to the membrane concomitantly removed the template RNA. The blot was prehybridised for 15 min in GOOP buffer (Appendix 1) and then probed with a radiolabeled PCR fragment which was complementary to the neurotoxin region under study (figure 3.8).

Reactions using either 3237 and 2677 oligonucleotide primers, with total RNA from *C. botulinum* NCTC 2916 or *C. botulinum* type B Danish produced virtually identical bands which occurred close to the ribosome binding site of the type B neurotoxin sequence (figure 3.9). The position of the one of these bands was approximately 5 bp in front of the proposed ribosome binding site which directly coincided with the result obtained with the standard primer extension method (figure 3.7).

3.5 Discussion.

3.5.1 A silent neurotoxin gene.

On the basis of nucleotide and predicted amino acid sequence similarities, *C. botulinum* NCTC 2916 would appear to possess the remnants of a neurotoxin gene which is most closely related to the proteolytic type B neurotoxin gene (Whelan *et al.*, 1992), classified as a silent type B neurotoxin gene (BoNT/B^o) in this study. This gene is in addition to the type A neurotoxin gene which had been previously characterised in this strain (Thompson *et al.*, 1990).

Although clostridia have been isolated that produce dual combinations of neurotoxin serotypes A:f (Gimenez *et al.*, 1970), B:f (Hatheway *et al.*, 1989) and A:B (Poumeyrol

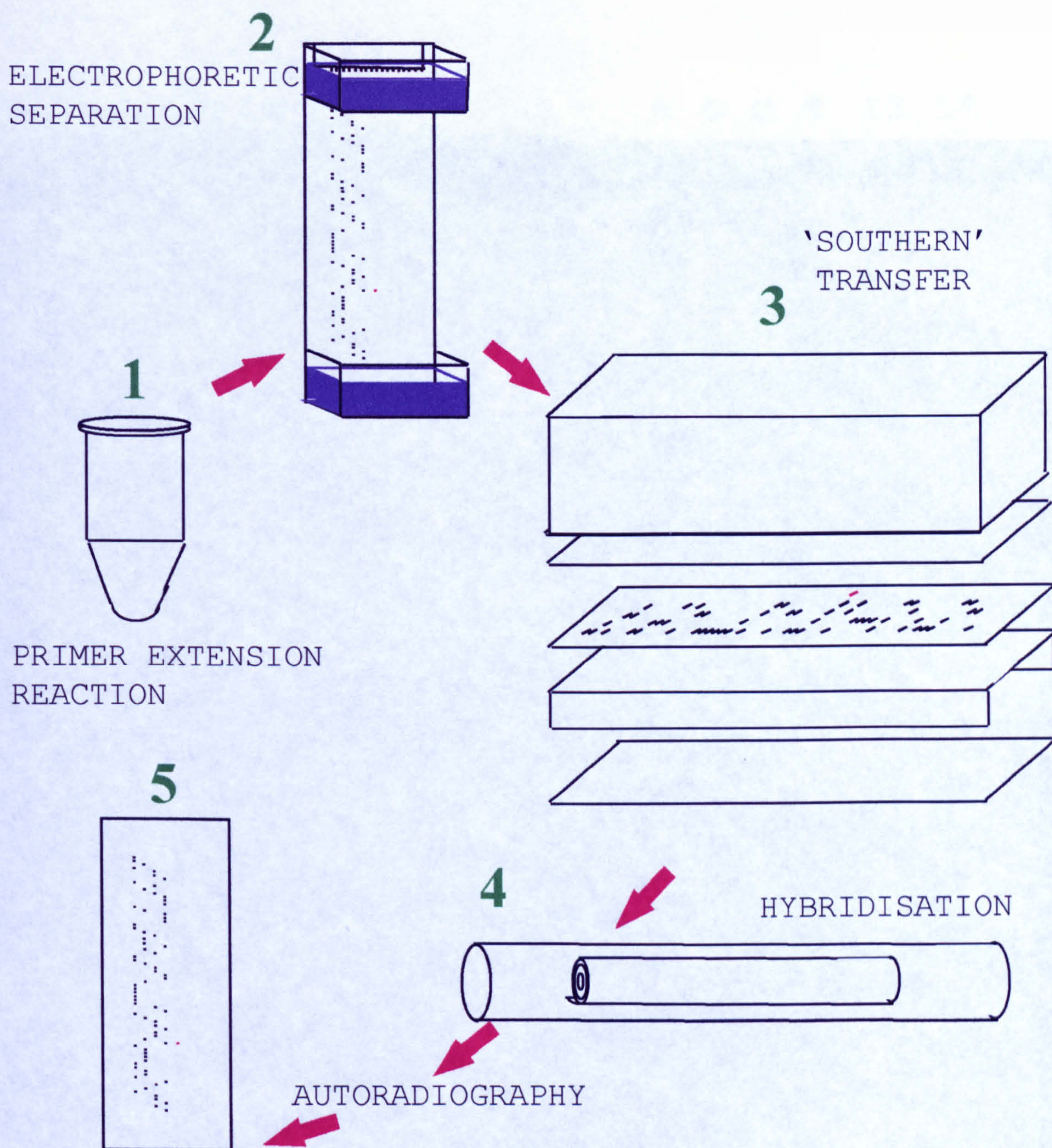


Figure 3.8 Development of a primer extension blotting technique.

To aid the identification of authentic primer extension products, an adaption to the primer extension method was made (section 3.4.3). In essence, primer extension products were radio-labeled after *in vitro* synthesis and electrophoretic separation, instead of starting with a radiolabeled primer in the cDNA synthesis reaction. 1) A primer extension reaction was performed with an unlabeled oligonucleotide primer. 2) The primer extension product (in red) were separated on a polyacrylamide gel, beside a control sequencing reaction (black). 3) The cDNA primer extension product and DNA control fragments were transferred to a Nylon membrane using a standard Southern blotting technique. 4) The membrane was then probed with a radiolabeled DNA fragment in a hybridisation solution which anneals to both primer extension products and DNA sequencing control. 5) The results were detected by conventional autoradiography.

A G C T L1 L2

A G C T L3 L4

5'
A
T
A
T
A
A
A
C
T
T
A
A
A
A
T
A
A
G
G
A
G
G
A
G
A
A
T
A
T
T
T
A
T
G
3'

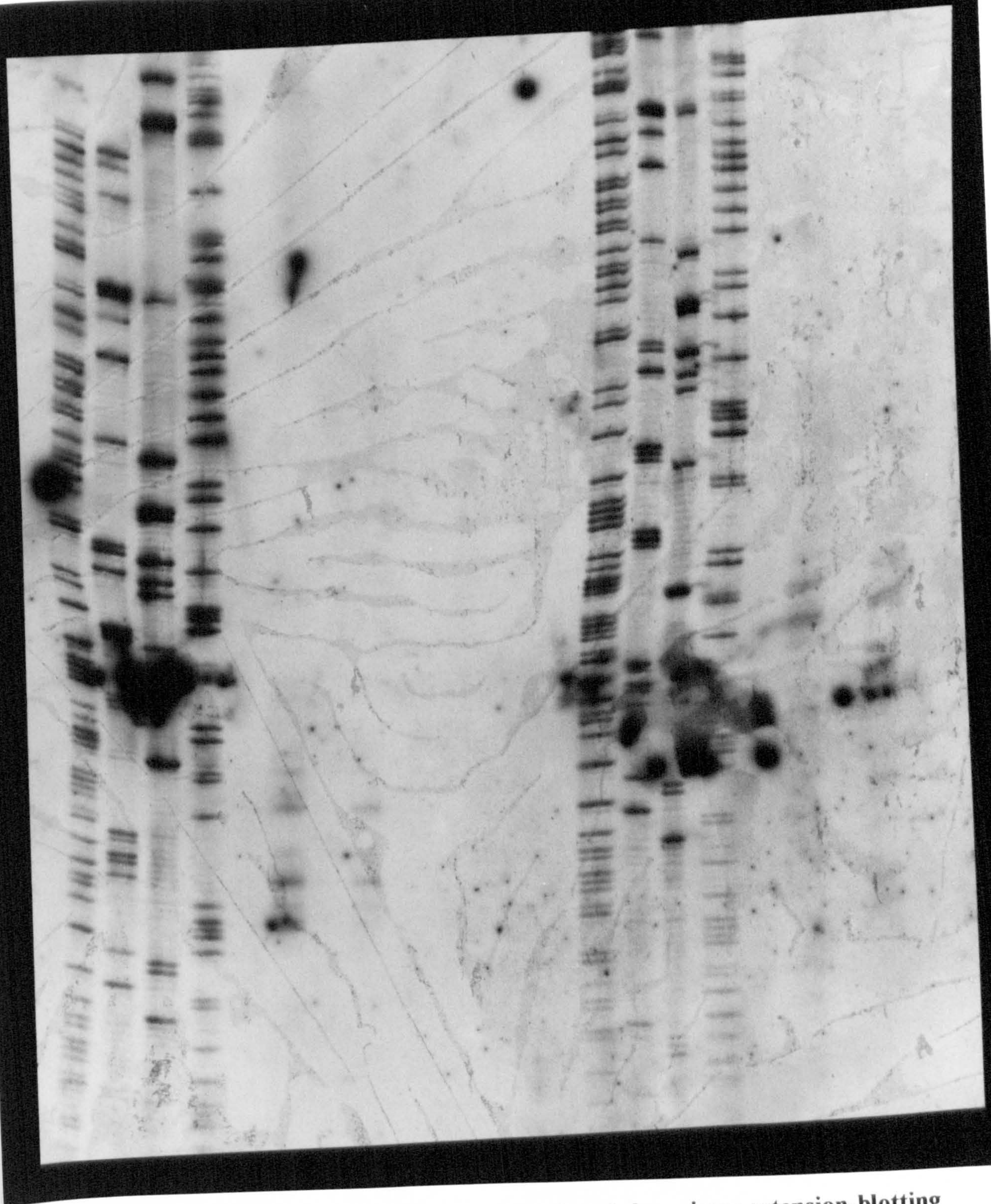


Figure 3.9 Detection of 5' mRNA ends by primer extension-blotting

Figure 3.9 Detection of 5' mRNA ends by primer extension-blotting

Transcripts were mapped by generating cDNA from unlabeled oligonucleotide primers using the method detailed in section 3.4.3. mRNA samples from *C. botulinum* NCTC 2916 and a proteolytic type B strain (Danish) were both extended with two separate primers 3237 and 2677 (Appendix V). Primer extension products were separated by denaturing gradient polyacrylamide gel electrophoresis beside the respective plasmid sequencing reaction (lanes AGCT). L1 and L3 are cDNA products generated from *C. botulinum* NCTC 2916 with primers 3237 and 2677 respectively whilst L2 and L4 are cDNA products generated from *C. botulinum* type B (strain Danish). The accompanying sequence is of the intergenic gap between the NTN_H gene and the silent BoNT/B where the stop codon of the NTN_H gene is blue, the base corresponding to the primer extension products in both standard and primer extension-blotting is tan, the ribosome binding site in red and the initiation codon of the silent type B neurotoxin gene in green.

et al., 1983, Gimenez *et al.*, 1984), the recent discovery of silent neurotoxin genes has added a new dimension to the genetics of this group of organisms. Indeed, the prevalence of silent type B genes within serotype A organisms has been reported to be relatively common (Franciosa *et al.*, 1994). This implies that the methods currently used to detect neurotoxins and their encoding genes are unsatisfactory. A protein detection approach will not detect a silent gene (irrespective of the reasons why the gene or part of the gene/toxin complex is not expressed) and a DNA based assay will not distinguish if the toxin gene(s) detected will be expressed. An additional concern with the presence of the silent toxin gene is that in theory the first 127 amino acids should be expressed and although the peptide may be degraded, it may equally persist in the cell. In light of the current pharmaceutical applications of clostridial neurotoxins, particularly type A, the presence of such peptides and associated components of the silent type B neurotoxin complex must be undesirable.

3.5.2 Genetic anomalies of the silent type B neurotoxin gene.

An interesting problem is posed when one considers the order in which sequence changes to the type B neurotoxin gene may have occurred. When comparing several neurotoxin genes (Elmore *et al.*, 1995) the deletion of several codons is not unusual. However, if the lysine and tyrosine residues were vital to the function of the neurotoxin, the organism would contain a particularly large protein which would be entirely redundant but still demand a lot of cellular resources to be expressed. This would also be the case with the C-terminal deletion mutation. A practical solution from the bacteria's point of view would be to minimise this futile energy loss by introducing a translational stop codon near the beginning of the gene and avoiding translating the large protein. Conversely the translational stop codon may have occurred first and the other mutations may simply be due to the lack of selective pressure for maintenance of the sequence.

During the course of this study another group published a partial sequence of a similar silent type B neurotoxin gene from the type A *C. botulinum* strain 667 (Hutson *et al.*, 1996). In addition to the genetic alterations described above an additional single base pair deletion was found (figure 3.10). This could suggest that the second frame-shift deletion in strain 667 was due to genetic drift within this sequence.

3.5.3 Interpretation of the Northern blotting

The Northern blots in figure 3.6a and 3.6b demonstrate that the silent type B neurotoxin gene in *C. botulinum* strain NCTC 2916 does have a mRNA transcript and that under the conditions isolated, the distribution of the toxin transcript appears to be virtually identical to the transcript detected in the proteolytic type B control organism, *C. botulinum* strain Danish. In fact, the distribution of the transcripts probed with type B genes appeared to resemble those of type A genes. In general, this appears to conform to a faint high molecular size transcript of at least 7.4 kb, which becomes more intense at approximately 1-3 kb. In between these sizes, a 'smear' can be observed. These observations may be explained by the physiological production of the neurotoxins. Studies by Bonventre & Kempe (1960), demonstrated that toxin production gradually increased, reaching a peak after approximately 24 h. In effect, toxin was being produced during exponential growth at a steadily increasing rate (section 1.3.3). The preparation of the clostridial cultures for purification of total RNA was designed to complement this phenomenon and this method had been successfully used to study the mRNA transcripts from the type A neurotoxin complex in strain NCTC 2916 (Henderson *et al.*, 1996). Thus, at the time of total RNA purification, full length transcript and partially degraded transcripts would have been present in all viable cells. The ratio of these transcripts would depend on the relative stability (half life) of the full length transcript. The deductions from the Northern blots could therefore be explained as a small amount of large full length transcript with shorter disintegrating transcripts increasing in number at 1-3 kb. The proposed size of the 'full length' mRNA suggested that the neurotoxin gene may be part of a polycistronic operon, possibly a bicistronic operon in combination with the preceding NTN_H gene (section 4.2.1.4). Attempts to detect neurotoxin specific transcripts from 'younger' cultures were unsuccessful which could be due to the low level of transcripts present. The toxin titer of the culture was not determined and, therefore, some degeneracy of toxin production can not be ruled out.

3.5.4 Interpretation of primer transcript analysis.

Studies on a number of other botulinal neurotoxin serotype transcripts have identified possible mRNA start points (Binz *et al.*, 1990; Hauser *et al.*, 1994; Henderson *et al.*,

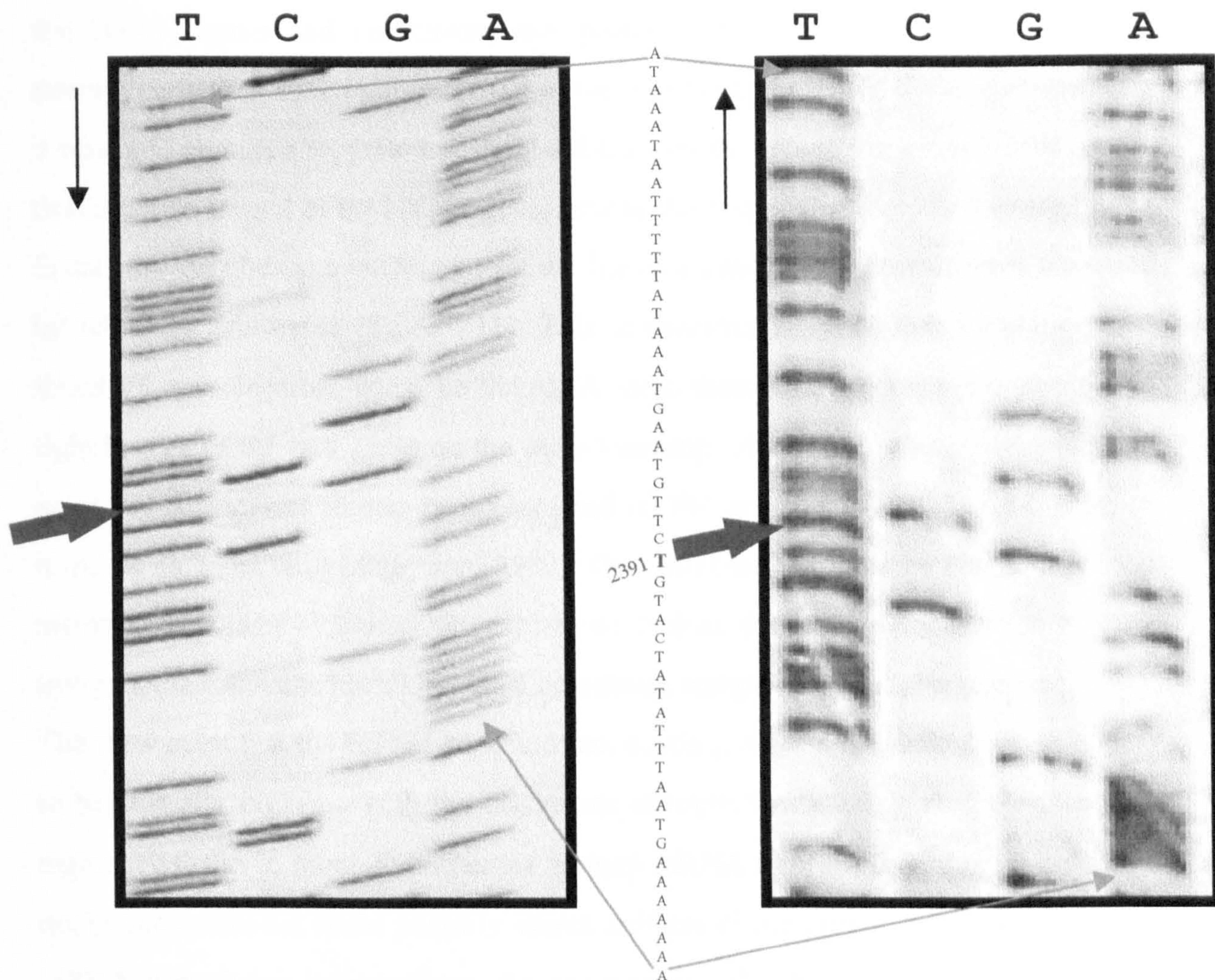


Figure 3.10 Nucleotide sequence discrepancy in the silent BoNT/B neurotoxin sequence

Comparison of the nucleotide sequences determined for the silent BoNT/B pseudogene in *C. botulinum* NCTC 2916 (this study) and the partial silent BoNT/B pseudogene in *C. botulinum* strain 667 (Hutson *et al.*, 1996) revealed a single discrepancy. In the sequence of Hutson (1996) a single thymidine base was deleted at position 2391. Autoradiographs of the equivalent region of sequence from this study were analysed (above) and the sequence determined in either direction (—→) was shown to contain a thymidine residue (in red) as was previously determined for the proteolytic type B (Whelan *et al.*, 1992) and non-proteolytic type B neurotoxin gene sequences (Hutson *et al.*, 1994).

1996). Through a combination of studying the length of the intergenic region between the NTN_H genes and neurotoxin, the position of the proposed transcripts and attempting to map a start point in front of the type B neurotoxin, it became evident that it was unlikely that a separate transcript did exist for the neurotoxin genes *per se*. In the first instance several of the NTN_H genes among the botulinal serotypes appeared to be fused with the ribosome binding site of the following neurotoxin gene or were separated by relatively few bases (figure 3.11). This organisation suggests that translation and, therefore, transcription, could be linked. A more theoretical argument supporting the tight linkage of the two genes on the same transcript stems from the observation that neurotoxin complexes contain neurotoxin and NTN_H proteins in the M and L complex forms at an equal ratio (Sugiyama, 1980). One mechanism to ensure that this ratio is maintained could be by linking the expression. Indeed, the relative organisation of these two genes is maintained in all botulinal neurotoxin complex operons characterised.

The view point that the NTN_H gene and neurotoxin gene could be linked would appear to be mutually exclusive with the occurrence of transcripts starting near the intergenic region. With this in mind, the attempts to map mRNA transcripts in this region would not be successful but could possibly detect artifacts of the primer extension process or mRNA degradation intermediates. An examination of a 500 bp section of sequence spanning the type B NTN_H-neurotoxin intergenic region did not appear to reveal an obvious secondary structure motif which could have been capable of prematurely terminating the reverse transcriptase. This suggests that if the bands which were obtained were not due to (multiple) promoters, they could be due to degradation products of a larger transcript.

A comparison of the transcripts mapped for the type A neurotoxin gene (Henderson *et al.*, 1996) with those determined for the silent and proteolytic type B neurotoxin genes (this study) revealed that a 'transcript' start point was detected 5' of the neurotoxin genes, which in both cases occurred at 5-6 bases in front of the respective gene's ribosome binding site. The positioning of a promoter so close to a gene is uncommon and was not found in the study of Binz (1990) who mapped a transcript start site for the type A neurotoxin to 118 bases upstream from the initiation codon. However, in that study the oligonucleotide primer bridged the type A neurotoxin ribosome binding site and would consequently have produced an equivalent transcript which was too

NTNH/A	5'-TGGTCATTTAAATAA-29-AAGAGG--8-ATG-3'	BoNT/A
NTNH/B	5'-TTAAAAGGGATATAA-10-AGGAGG--9-ATG-3'	BoNT/B
NTNH/C	5'-ATATTAGAAAGTTAG-----GAGA-10-ATG-3'	BoNT/C
NTNH/D	5'-ATATTAGAAAGTTAG-----GAGA-10-ATG-3'	BoNT/D
NTNH/E	5'-TTAAAAAATATATAA-10-AGGAGA--8-ATG-3'	BoNT/E
NTNH/F	5'-TTGAAAGAAGTATAG-----GGGG--8-ATG-3'	BoNT/F
NTNH/G	5'-TTAAAAAAGATATAG-70-AGGAGG--8-ATG-3'	BoNT/G

Figure 3.11 Alignment of the intergenic regions between botulinal neurotoxin serotypes

The intergenic region between the NTN gene and neurotoxin gene are aligned for the different neurotoxin serotypes so a rational prediction of translational coupling can be made. The last four codons of the NTN gene are shown, followed by the stop codon (green), ribosome binding site (red) and initiation codon of the neurotoxin. Nucleotide spacings are depicted by numbering and nucleotides shared between the stop codon and ribosome binding site are in blue. The sequence evidence suggests that the NTN gene and neurotoxin gene of serotypes C, D and F have tight translational coupling, those of B and E would also appear to be coupled whereas the sequence evidence appears to be less obvious for A and G. However, due to the conflicting evidence supporting the 5' positioning of mRNA transcripts (Henderson *et al.*, 1996, Binz *et al.*, 1990, Hauser *et al.*, 1994) the existence of a promoter positioned within the NTN coding sequence cannot be ruled out. On the other hand, the requirement of a BoNT-specific promoter would not be logical if the NTN and neurotoxin gene were transcriptionally and translationally coupled.

small to detect. A survey of the literature did describe a phenomenon where short polypurine tracts of mRNA were capable of attenuating mRNA degradation (Hue *et al.*, 1995; Agaisse & Lereclus, 1996). These sequences, which closely resemble ribosome binding sites, are thought to protect the transcript from 5'→3' ribonucleolytic degradation via binding of a ribosome to produce a pseudo-transcriptional start site a few bases in front of the polypurine tract (Agaisse & Lereclus, 1996). An obstacle to this hypothesis is that little evidence for a 5'→3' ribonuclease exists, at least in *E. coli*. However, recent studies in *B. subtilis* suggest that this activity may exist (DiMari & Bechhofer., 1993) and that 5'-proximal structures which sequester the 5' terminus of an mRNA can increase a transcripts half life. The explanation proposed is that ribonucleases require either a 5' or 3' end from which to initiate degradation (figure 3.12). Among the mechanisms which appear to protect the 5' end of a transcript from ribonucleolytic degradation, ribosome binding sites appear to feature prominently (table 3.2). Studies on *E. coli lacZ* transcripts provide evidence that the efficiency of the SD is a strong determinant of transcript stability and that the SD sequence may effect this role independent of its role in the initiation of translation (Agaisse & Lereclus, 1996; Wagner *et al.*, 1994).

The prevalence of 'false' transcripts due to the proposed phenomenon is not widely reported, possibly due to the heterogeneous nature of DNA sequences to which promoter elements can be assigned. A survey by Agaisse & Lereclus (1996) did detect a number of transcripts which appeared to initiate shortly before a 'SD type sequence'. In addition the SD sequence of the *B. subtilis* SP82 bacteriophage early RNA has been demonstrated to perpetuate transcript half life (Hue *et al.*, 1995).

Assuming that this transcript interpretation is correct, what purpose could such a stabilising element serve in the inter-cistronic space between NTNH and BoNT genes? Without further practical experimentation, this question cannot be answered. However, several plausible ideas can be suggested. Ehretsmann (1992) noted that in general, mRNA degradation is rapid, where the majority of transcripts in procaryotes such as *E. coli* have a half life of 2-3 minutes. Thus, in a long polycistronic transcript, mRNA degradation is sufficiently rapid that the 5' end begins to be degraded before the 3' end is synthesised. Thus, differences in the stability of separate genes of a polycistronic transcript may be required to maintain proper co-ordination of the expressed proteins

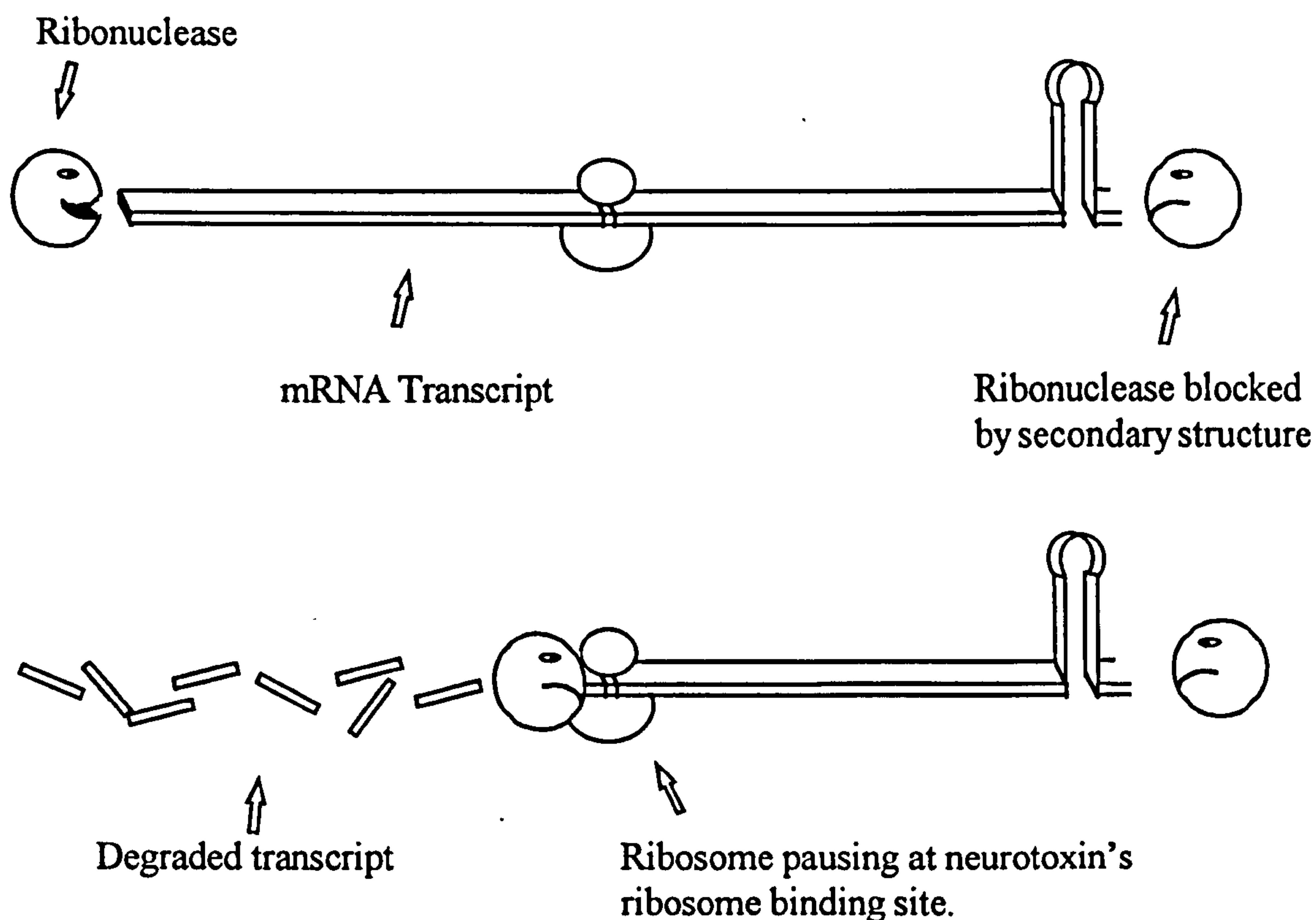


Figure 3.12 Hypothetical mechanism of NTNH-Neurotoxin transcript degradation

Ribonucleolytic degradation of the NTNH-neurotoxin transcript may occur by the action of a 5'-3' ribonuclease or by the 5' binding of a 3'-5' ribonuclease (section 3.5.4). The 3' end of the transcript is probably protected from degradation by an inverted repeat which may also act as a Rho-independent transcriptional terminator. Degradation which has initiated at the 5' end would be blocked by a ribosome pausing at the ribosome binding site of the neurotoxin transcript. As the half life of the neurotoxin transcript would be longer than that of the preceding NTNH section, the new 5' end of the neurotoxin transcript would be preferentially detected by techniques such as primer extension analysis (adapted from Bechofer, 1991)

Assuming that an 8 kb transcript of this nature is relatively long, protection from 5' binding endoribonucleases may be of particular relevance. Another hypothesis could be that translation of the NTNH and BoNT/B genes by a ribosomal complex are intimately linked in such a manner that the two newly synthesised peptides are able to efficiently bind during toxin complex assembly (figure 3.13). If this were the case, a brief pause between the genes may be necessary to enable a post translational modification step such as peptidase processing of the NTNH protein to occur. Theoretical justification of this idea is supported by the observation that NTNH and BoNT proteins are naturally found processed and complexed. The sequestration of the neurotoxin may be vital to avoid a deleterious intracellular action of the neurotoxin which could be similar to the activity observed with transaminases in eucaryotic cells.

ORGANISM	LOCUS	MECHANISM	COMMENT	REFERENCE
<i>B. thuringiensis</i>	<i>cryIIIA</i> toxin gene	Shine-Dalgarno sequence	interaction of a S-D sequence in the mRNA leader sequence with a ribosome protects from 5'→3' ribonucleolytic degradation.	Agaisse & Lereclus (1996)
<i>B. thuringiensis</i>	<i>cryIIIB</i> toxin gene	Shine-Dalgarno sequence	suggested by Agaisse & Lereclus. (1996)	Donavon <i>et al.</i> (1992)
<i>B. thuringiensis</i>	<i>cryIIIB2</i> toxin gene	Shine-Dalgarno sequence	suggested by Agaisse & Lereclus. (1996)	Donavon <i>et al.</i> (1992)
<i>B. brevis</i>	<i>cwp</i>	Shine-Dalgarno sequence	suggested by Agaisse & Lereclus. (1996)	Adachi <i>et al.</i> (1989)
<i>L. monocytogenes</i>	<i>inlAB</i>	Shine-Dalgarno sequence	suggested by Agaisse & Lereclus. (1996)	Dramsi <i>et al.</i> (1993)
<i>B. subtilis</i>	early RNA of bacteriophage SP82	Shine-Dalgarno like sequence	proposed interaction with the 3' end of 16 S rRNA	Hue <i>et al.</i> (1995)
<i>B. subtilis</i>	<i>ermC</i>	ribosome interaction with mRNA	erythromycin induced stalling of a ribosome	Bechhofer & Zen (1989)
<i>S aureus</i>	<i>ermA</i>	ribosome interaction with mRNA	erythromycin induced stalling of a ribosome	Sandler & Weisblum, (1989)

Table 3.2 Examples of mRNA stability influenced by Shine-Dalgarno type sequences

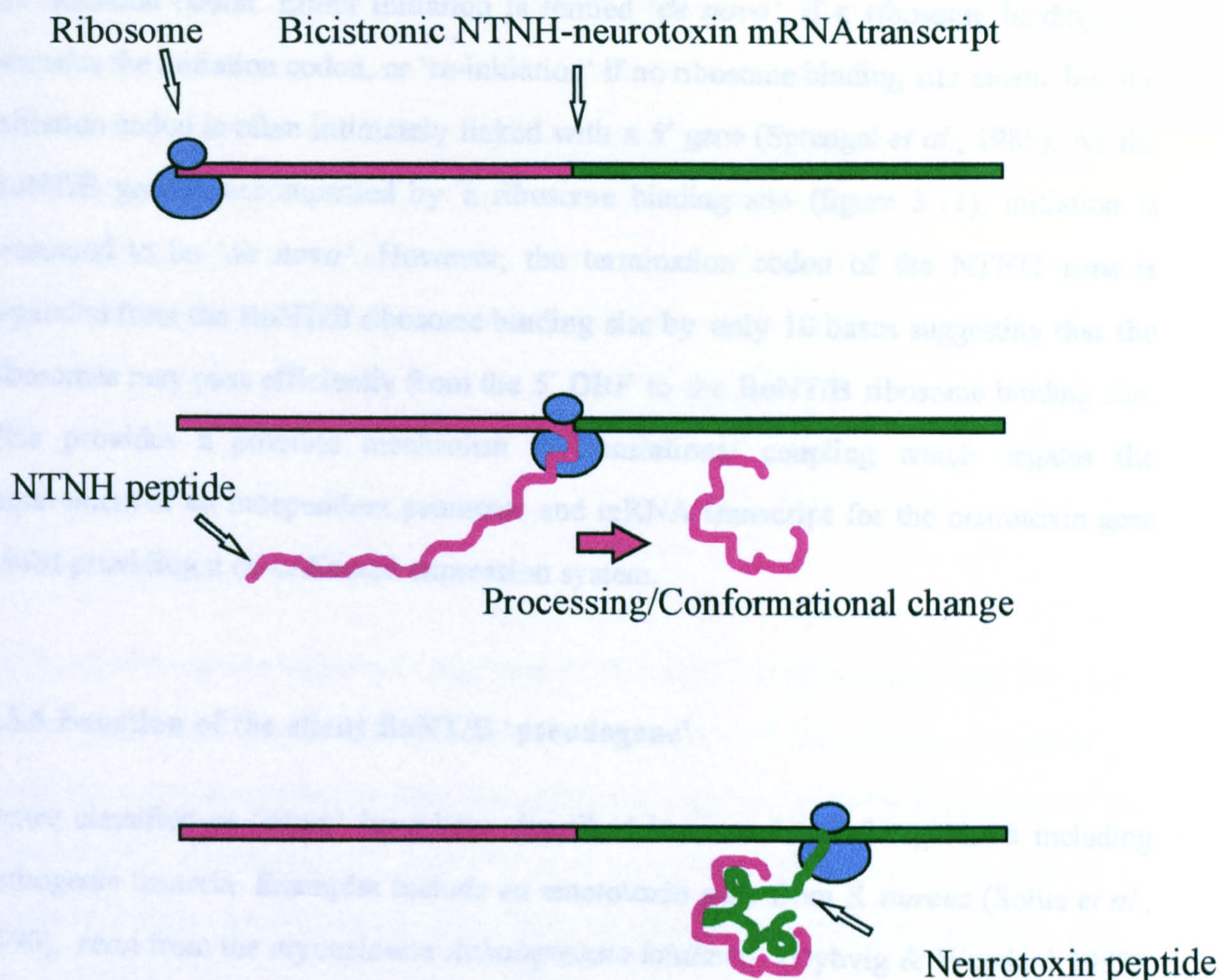


Figure 3.13 Hypothetical function of the stalled ribosome.

A possible function of the ribosome stalling at the neurotoxin's SD sequence could be that a short delay may be required between the translation of the NTNH gene and the neurotoxin gene. This could be essential if precise sequestration of the neurotoxin by the NTNH protein was important to protect the cells from proteolysis or the formation of transmembrane pores by the neurotoxin (section 3.5.4).

3.5.5 Translational coupling of the NTNH and BoNT/B transcripts?

The efficiency of translational initiation is partly determined by the sequences preceding the initiation codon. Either initiation is termed '*de novo*', if a ribosome binding site precedes the initiation codon, or 're-initiation' if no ribosome binding site exists, but the initiation codon is often intimately linked with a 5' gene (Sprengel *et al.*, 1985). As the BoNT/B gene is accompanied by a ribosome binding site (figure 3.11), initiation is presumed to be '*de novo*'. However, the termination codon of the NTNH gene is separated from the BoNT/B ribosome binding site by only 10 bases suggesting that the ribosomes may pass efficiently from the 5' ORF to the BoNT/B ribosome binding site. This provides a possible mechanism of translational coupling which negates the requirement of an independent promoter and mRNA transcript for the neurotoxin gene whilst providing a co-ordinated expression system.

3.5.6 Function of the silent BoNT/B 'pseudogene'

Genes classified as 'silent' have been described in a variety of organisms including pathogenic bacteria. Examples include an enterotoxin gene from *S. aureus* (Soltis *et al.*, 1990), *recA* from the mycoplasma *Acholeplasma laidlawii* (Dybvig & Woodard 1992), *ilvG* genes from *E. coli* K12 and *Salmonella typhimurium* LT2 (Burns *et al.*, 1995) and pilin genes from *Neisseria gonorrhoeae* (Gibbs *et al.*, 1989). However, interpreting the biological significance of such 'silent' genes is complex. The most simple scenario is proposed for the *ilvG* haemolysin of *E. coli* K12 and *S. typhimurium* LT2. Here the nonsense mutations are thought to have been acquired since entering the laboratory, a phenomenon also proposed for other laboratory isolates (Burns *et al.*, 1995). Such an explanation is unlikely for the mutations occurring in the BoNT/B genes in group I *C. botulinum* strains as their occurrence appears to be widespread in a number of separate isolates (Franciosa *et al.*, 1994). An alternative hypothesis could be that the pseudogene sequences have an important role in the pathogenesis of neurotoxigenic clostridia, which may be maintained as a transiently redundant gene pool. This type of explanation is illustrated in *N. gonorrhoeae* where novel pilin genes are assembled from a larger

repertoire of partial gene sequences (Gibbs *et al* 1989). This type of mechanism is a versatile method of increasing or maintaining virulence, which is used by a number of pathogens, including trypanosomes (Roth *et al.*, 1989). A neurotoxin gene has been isolated which appears to consist of a type C and type D chimera (Moriishi *et al.*, 1996). Although the process by which this gene evolved is unknown, it must be noted that the majority of neurotoxigenic clostridia have been isolated from clinical cases. Thus, only functional genes would be isolated, and therefore we may only see a fraction of the clostridia which do possess active neurotoxins and not those which maintain a library of 'transiently redundant' neurotoxin sequences from which chimeras could be constructed.

3.5.7 Reversion to an active neurotoxin?

From an analysis of the nucleotide sequence of this BoNT/B silent gene, two principle mutations appear responsible for the discontinuity of the coding sequence; a substitution at nucleotide 381 to form a translational stop codon and a single nucleotide deletion at 2881. Assuming that rectification of only these defects would be necessary to produce a viable neurotoxin, the possibility of this gene returning to an active form must be feasible. This, however, does not seem to have been the case for the partially characterised AB strain I.P. 7212 (Fujinaga *et al.*, 1995) which is virtually identical to the type B neurotoxin from strain Danish.

In some instances a translational stop codon does not present a total barrier to translation (Gozalbo & Hohmann, 1989). Indeed, the identity of the bases which reside next to a stop codon can to some extent modulate the efficiency to which the stop codon can be suppressed (Feinstein & Altman, 1978). In *E. coli*, the effect of 'codon context' was shown to be due to the 5' and 3' codons surrounding the stop codon (Bjornsson & Isaksson, 1993). Purine bases which are under represented at 'true' termination codons (Brown *et al.*, 1990) were able to enhance leaky translational readthrough which was further increased by mutations in specific ribosomal proteins. If translational readthrough is also feasible in clostridia, a possibility exists that a low level of

expression, or base changes which lead to a conducive codon context around the fortuitous stop codon will enable a low level of expression of the BoNT/B gene. As clostridial neurotoxins are particularly potent, this level of expression may be significant.

3.5.8 Why do type A group I *C. botulinum* strains commonly contain type B pseudogenes?

A number of biologically important proteins depend on the aberrant suppression of stop codon translational termination to create a greater protein repertoire (Ryoji *et al.*, 1983) but this does not seem to be likely in this example. Instead, a rational explanation could be that other components of the neurotoxin complex which are encoded at the silent BoNT/B locus do provide an advantage to the organism,. e.g., the NTNH or haemagglutinin proteins. If this were the case, the BoNT/B neurotoxin may have become an unnecessary burden to an organism which already contained an efficient neurotoxin. Expression of the NTNH gene may depend on the stability afforded by the 3' BoNT/B transcript, thus a nonsense stop codon near the 5' end of the BoNT/B would avoid expression of the BoNT/B neurotoxin but still maintain NTNH gene expression.

CHAPTER 4

MOLECULAR CHARACTERISATION OF THE NEUROTOXIN OPERON FOUND IN ASSOCIATION WITH THE TYPE B NEUROTOXIN GENE IN *C. botulinum* NCTC 2916.

4.1 Introduction

Genetic evidence accumulated whilst cloning and characterising the botulinal neurotoxins serotypes A-G (Elmore *et al.*, 1995) demonstrated that at least one additional component of the neurotoxin complex was consistently found to reside 5' to neurotoxin genes. Initially, an ORF was found in the same orientation and in close proximity to the neurotoxin genes of serotypes A, C, D and E (Binz *et al.*, 1990a; Kimura *et al.*, 1990; Binz *et al.*, 1990b; Fujii *et al.*, 1990). Alignment of the predicted C-terminal sequence of the ORF from the different serotypes demonstrated a high degree of similarity. A further report established that the ORF which was predicted upstream of the type C neurotoxin gene was indeed a component of the neurotoxin complex (Tsuzuki *et al.*, 1992). In addition, a clone whose gene product possessed haemagglutinin activity was also located 5' to the type C neurotoxin locus (Tsuzuki *et al.*, 1990). This suggested that the haemagglutinin activity associated with other neurotoxin complex serotypes (Somers & DasGupta, 1991) could also be genetically linked to the respective neurotoxin gene. With these facts in mind, a strategy was devised to identify additional genetic components associated with the silent type B neurotoxin gene from *C. botulinum* NCTC 2916.

4.2 Results

4.2.1 Characterisation of the silent type B neurotoxin complex locus

4.2.1.1 Evidence of neurotoxin complex genes accompanying the silent type B neurotoxin gene.

Initial data supporting the existence of additional genes 5' to the silent type B neurotoxin gene surfaced whilst characterising the silent neurotoxin gene (section 3.3.2).

Southern blotting with DNA fragments from the NTNH/A gene generated data incompatible with the type A neurotoxin locus alone (Dr S. M. Whelan pers. comm.) In addition, the preliminary sequence data from clone pSB1 supported this view (section 3.3.2), but also confirmed previous reports that the components of type A and B neurotoxin complexes were closely related (Somers & DasGupta, 1991). Attempts to clone DNA fragments 5' to the silent type B neurotoxin were complicated by the prediction that the type A and type B NTNH and haemagglutinin genes were very similar (Somers & DasGupta, 1991). Type B specific NTNH and haemagglutinin genes were reliably cloned by targeting a DNA fragment contiguous with part of the type B neurotoxin gene (section 3.3.5).

4.2.1.2 Cloning strategy

A PCR fragment of the type B NTNH gene (section 3.3.3) which differed from the type A NTNH gene (Henderson *et al.*, 1996) was excised as a *Bgl*III-*Stu*I restriction fragment. This was radio-labeled and used in Southern blotting experiments to create a restriction map of the contiguous sequence (figure 4.1 and figures 4.7a, 4.7b). Accordingly, a 6.5 kb *Xba*I restriction fragment was targeted for cloning into an *E. coli* vector pMTL22 (Chambers *et al.*, 1988). The entire nucleotide sequence of the cloned fragment, pSB1 was determined on both strands by the method of Sanger (section 2.2.11.4). Forward universal and reverse universal oligonucleotide primers were used to obtain initial sequence data, from which successive oligonucleotides were designed to 'walk' through the clones. Individual sequences were compiled into a single contig using the DNASTAR software package 'Seqman'. The complete nucleotide sequence was translated in all six

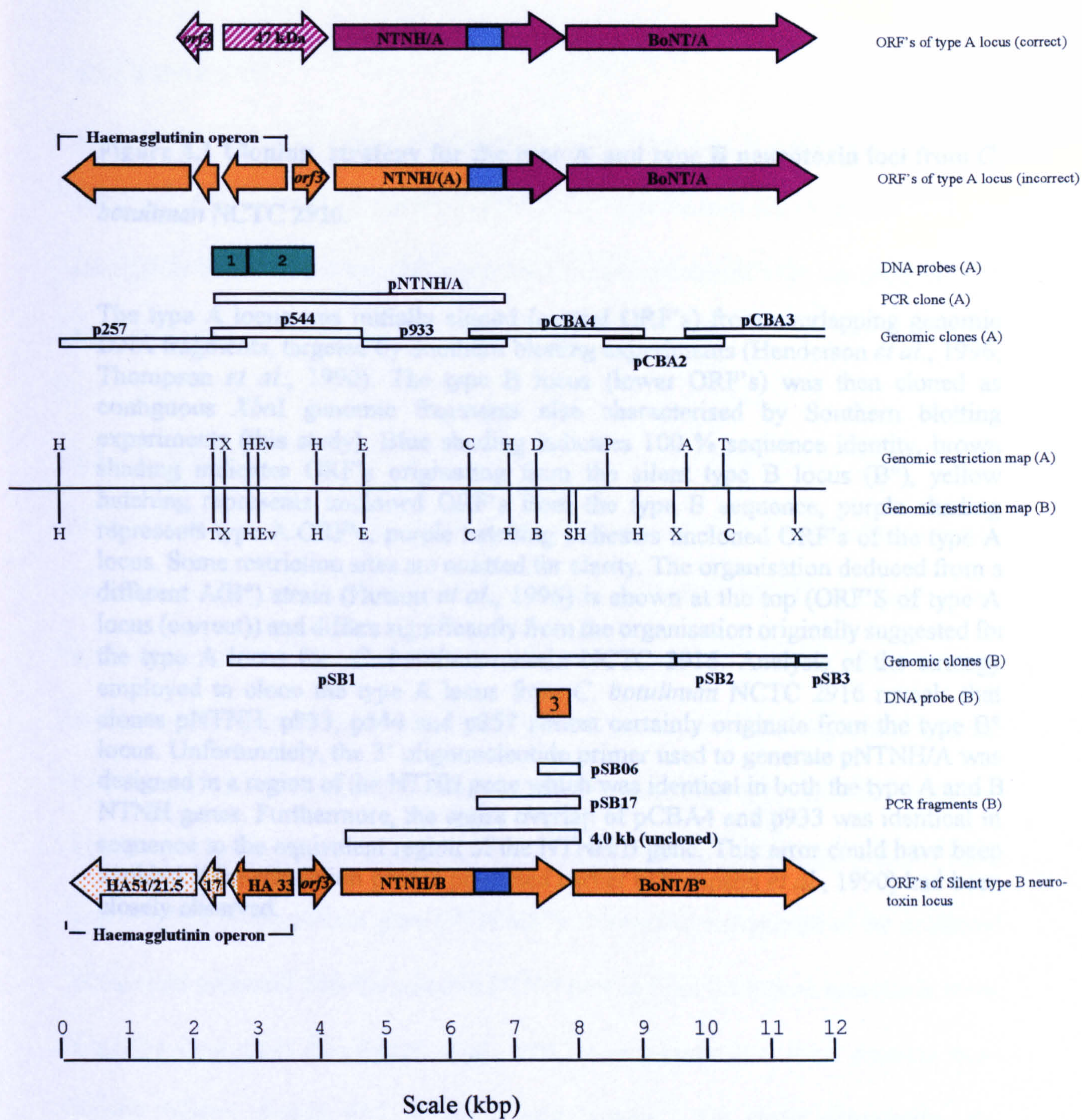


Figure 4.1 Cloning strategy for the type A and type B neurotoxin loci from *C. botulinum* NCTC 2916.

Figure 4.1 Cloning strategy for the type A and type B neurotoxin loci from *C. botulinum* NCTC 2916.

The type A locus was initially cloned (central ORF's) from overlapping genomic DNA fragments, targeted by Southern blotting experiments (Henderson *et al.*, 1996; Thompson *et al.*, 1990). The type B locus (lower ORF's) was then cloned as contiguous *Xba*I genomic fragments also characterised by Southern blotting experiments (this study). Blue shading indicates 100 % sequence identity, brown shading indicates ORF's originating from the silent type B locus (B°), yellow hatching represents uncloned ORF's from the type B sequence, purple shading represents type A ORF's, purple hatching indicates uncloned ORF's of the type A locus. Some restriction sites are omitted for clarity. The organisation deduced from a different A(B°) strain (Hutson *et al.*, 1996) is shown at the top (ORF'S of type A locus (correct)) and differs significantly from the organisation originally suggested for the type A locus for *C. botulinum* strain NCTC 2916. Analysis of the strategy employed to clone the type A locus from *C. botulinum* NCTC 2916 reveals that clones pNTNH, p933, p544 and p257 almost certainly originate from the type B° locus. Unfortunately, the 3' oligonucleotide primer used to generate pNTNH/A was designed in a region of the NTNH gene which was identical in both the type A and B NTNH genes. Furthermore, the entire overlap of pCBA4 and p933 was identical in sequence to the equivalent region of the NTNH/B gene. This error could have been avoided if the restriction profile predicted earlier (Thompson *et al.*, 1990) had been closely observed.

possible reading frames. Two complete ORF's and two partial ORF's were detected (figure 4.1).

4.2.1.3 ORF 1

The 5' end of a gene previously characterised as the silent type B neurotoxin was found at the end of this genomic fragment (section 3.3.6). This verified that the genes which are contiguous within the 6.5 kb *Xba*I clone are naturally associated with the silent type B neurotoxin gene.

4.2.1.4 ORF 2

ORF 2 coded for a protein of 1197 amino acids starting with a methionine initiation codon and ending with a ochre termination codon (Appendix III). The primary amino acid sequence of ORF 2 showed similarity with the N-terminal sequence data determined by protein micro-sequencing for the 'nontoxic-nonhaemagglutinin' protein (NTNH) isolated from type C (Tsuzuki *et al.*, 1992) and type E (Somers & DasGupta, 1991) botulinal neurotoxin complexes. This strongly indicated that ORF 2 coded for a nontoxic-nonhaemagglutinin gene (NTNH/B°). Comparative alignment of the predicted amino acid sequence with the putative NTNH protein from the type A neurotoxin locus from the same strain (Henderson *et al.*, 1996) demonstrated that these proteins were closely related, sharing 93.7 % amino acid identity. On closer examination, the alignment of the putative type A and B NTNH genes appeared to show regions of almost perfect identity, particularly in the N-terminal 1021 amino acids where only a single base change was detected. However, the 177 C-terminal amino acids differed significantly from the corresponding NTNH/A gene and were virtually identical to the

C-terminal region of the NTN^H/B gene from the proteolytic type B neurotoxin locus (Whelan *et al.*, 1992). In an attempt to judge if the putative NTN^H/B^o was representative of a type B neurotoxic strain, the sequence of another NTN^H/B gene was aligned with the NTN^H genes from *C. botulinum* NCTC 2916 (Appendix IV). The comparison demonstrated that NTN^H/B^o possessed greater similarity with the N-terminal 1021 amino acids of NTN^H/A from the same strain and not with the published type B strain. However, in general it appeared that the degree of similarity among NTN^H genes from various serotypes appeared to alternate inconsistently along the length of the genes.

Contemporary studies have demonstrated that a representative NTN^H gene is present in all other botulinal serotypes investigated and that its genetic orientation is conserved with respect to the accompanying neurotoxin gene (East *et al.*, 1994). A comparative alignment of the predicted primary amino acid sequences available for serotypes A-F (Appendix IV) demonstrates that these proteins are well conserved. In addition, they contain a number of N-terminal amino acid motifs which are also present in the N-terminus of all currently characterised clostridial neurotoxins (Elmore *et al.*, 1995). Both the composition and spatial orientation of these genes with their respective neurotoxin adds further weight to the assumption that they are homologs.

4.2.1.5 ORF 3

In a region spanning 925 bp 5' to NTN^H/B^o, a second ORF of 178 amino acids was identified. This was encoded on the same strand as NTN^H/B^o and would be predicted to be transcribed in the same orientation as both the NTN^H/B^o and neurotoxin gene. The putative methionine initiation codon was preceded by a polypurine sequence resembling

a Shine-Dalgarno sequence (Shine & Dalgarno, 1974). Although the mass of this predicted protein (21.65 kDa) was close in size to components of the type A neurotoxin protein complex (Somers & DasGupta, 1991), no sequence similarities could be detected. ORF 3 did not appear to encode a protein encountered in the neurotoxin complex. A search of the entire Swiss-Prot data base (release 100) revealed a low degree of amino acid similarity with a number of proteins. A 29 % identity over a 61 amino acid overlap occurred between the ORF 3 polypeptide and MsmR, a positive regulatory effector controlling the multiple sugar metabolism (msm) system in *Streptococcus mutans* (Russell *et al.*, 1992). From a critical examination of this alignment, the majority of the similarity was centered in a section which was described as a helix-turn-helix motif in MsmR (Russell *et al.*, 1992). The equivalent section of the ORF 3 polypeptide was evaluated for helix-turn-helix potential using the method of Dodd & Egan (1990) and was predicted to contain this motif (figure 4.2).

Application of this technique to other proteins retrieved in the database search included UviA, (a protein implicated in the UV-irradiation response of *C. perfringens*; Garnier & Cole, 1988), and ORFtxe1 (present on the *C. difficile* toxigenic element; Hammond & Johnson, 1995), also detected helix-turn-helix motifs (figure 4.2). However, unlike MsmR, both UviA and ORFtxe1 were of similar length to the ORF 3 protein and shared similarity along their entire length (figure 4.3).

Among the other proteins which were identified in the database search were a variety of transcriptional factors, however, the alignments did not provide compelling evidence of an evolutionary or functional link. Several sigma factors displayed a low level of sequence similarity with the ORF 3 polypeptide, UviA and the ORFtxe1 polypeptide.

Among these were, SigG (12.1 %) from *C. acetobutylicum* (Sauer *et al.*, 1994), SigG

AMINO ACID POSITION IN PUTATIVE H-T-H																						
Amino Acid	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A	-125	-194	-84	70	36	54	238	-15	77	26	-194	-194	-56	-84	14	77	-56	-56	-56	48	-195	36
C	-64	-64	-63	-63	-64	-64	-64	-64	-64	47	47	-63	-63	-64	-64	-64	-64	-64	-64	-63	-64	47
D	-156	-154	-156	-154	109	-156	-156	109	-154	-156	6	-156	-154	-85	-156	-156	-156	-156	-154	-154	-156	-85
E	-31	-9	-171	70	156	-171	-171	107	50	-60	-60	-171	-60	78	86	-171	-171	-101	-171	-170	86	9
F	10	-130	10	-130	-130	10	-130	-129	-130	-102	-130	-130	-130	-130	-129	-130	-130	-129	-129	-129	180	-130
G	30	5	-190	-51	-191	-191	18	-191	-191	-191	202	-191	-10	-191	5	-190	-191	-80	-190	-190	-191	-51
H	62	33	-76	-76	-7	-78	-78	33	-7	-78	84	-78	33	33	-78	-7	-78	-7	62	84	-78	-7
I	75	-156	101	-45	-86	116	-156	-16	65	-16	-156	128	-156	-86	-156	-155	188	-155	-16	53	122	-155
K	-31	-31	10	70	79	-170	-170	64	70	-171	-9	-100	-100	25	-100	-170	-171	-9	38	-31	-9	101
L	66	-212	72	-213	-212	144	-213	-102	37	132	-213	97	-213	-142	-212	-212	97	-212	-212	37	88	-213
M	122	-74	-3	-73	-73	-73	-74	-74	88	122	-73	258	-74	-74	-3	-74	-3	-74	-74	-3	-74	-73
N	-137	72	-137	136	-137	-137	-137	-67	-136	-136	128	-137	72	-136	2	-67	-137	2	84	-137	-137	104
P	-156	23	-157	-156	-157	-157	-157	-157	-157	-157	-157	-157	-46	101	39	-157	-157	-157	-157	-157	-157	-46
Q	-60	-130	175	90	110	-131	-131	90	78	-131	-131	-60	-130	154	65	119	-131	-20	119	31	-20	90
R	65	76	110	65	7	-155	-154	123	76	-155	-154	-155	-154	129	54	40	-155	129	179	-45	-155	123
S	-118	96	-188	21	-48	-187	-8	-187	-118	-118	-77	-188	174	-187	135	-26	-188	150	-77	-188	-187	-26
T	11	149	-59	80	-169	-8	-170	-99	-99	-30	-170	-8	131	-30	-59	-198	-170	-30	-169	-170	-30	-59
V	17	-67	-177	-177	-108	100	-178	-178	-108	71	-178	160	-178	-16	-67	-67	169	-178	17	31	17	-178
W	44	-26	-26	-26	-26	-26	-26	-26	-26	-25	-26	-25	-26	-26	-26	-26	-26	-26	44	279	-26	-26
Y	-40	-110	30	1	-110	-109	-110	-110	-40	30	-110	-109	-109	-40	-110	-40	-110	162	52	86	-110	-110

Table 4.1 Weight matrix for prediction of helix-turn-helix motifs.

Prediction of helix-turn-helix motifs in ORF 3₁₄₁₋₁₆₃, uviA₁₅₄₋₁₇₆ and txeR₁₄₅₋₁₆₇.

ORF3	I	K	E	I	D	I	A	K	K	L	N	I	S	R	Q	S	V	Y	K	N	K	I
	[75]	[-31]	[-171]	[-45]	[109]	[116]	[238]	[94]	[70]	[132]	[128]	[128]	[174]	[129]	[65]	[-26]	[169]	[162]	[38]	[-137]	[-9]	[-155]
UviA	K	S	E	V	E	I	A	E	I	M	G	T	S	R	Q	W	I	N	R	I	K	N
	[-31]	[96]	[-171]	[-177]	[156]	[116]	[238]	[107]	[65]	[122]	[202]	[-8]	[174]	[129]	[65]	[-26]	[188]	[2]	[179]	[53]	[-9]	[104]
TxeR	Y	S	D	R	E	I	S	V	K	L	K	I	S	R	Q	A	V	N	K	A	K	N
	[-40]	[96]	[-156]	[65]	[156]	[116]	[-8]	[-178]	[70]	[132]	[-9]	[128]	[174]	[129]	[65]	[77]	[169]	[2]	[38]	[46]	[-9]	[104]

(H-T-H)	CUMULATIVE SCORE	SCORE CLASS	H-T-H PROBABILITY
ORF 3	1253	3.5	71 %
UviA	1574	4.5	100 %
TxeR	1167	3.2	>50 %

Table 4.2 Calculation of helix-turn-helix probability.

Figure 4.2 Prediction of H-T-H motifs

The probability of H-T-H potential was estimated by the ‘Weight matrix method’ of Dodd & Egan, (1990). Individual amino acid scores relating to their position in the H-T-H (table 4.3.1) were calculated to give a cumulative score (cs) and entered into an equation to derive a ‘score class’ figure (Dodd & Egan, 1990) i.e. $[cs - 238.1] / 293.61 = \text{score class}$, from which a H-T-H probability was derived. Prediction of the H-T-H by this method relies on the reference set of H-T-H sequences to accurately represent all such motifs, therefore sequences which are not predicted to form this motif may do so but may not be fairly represented in the master reference set, an example is TrpR (Dodd & Egan, 1990). This method must be only used as a positive discriminator.

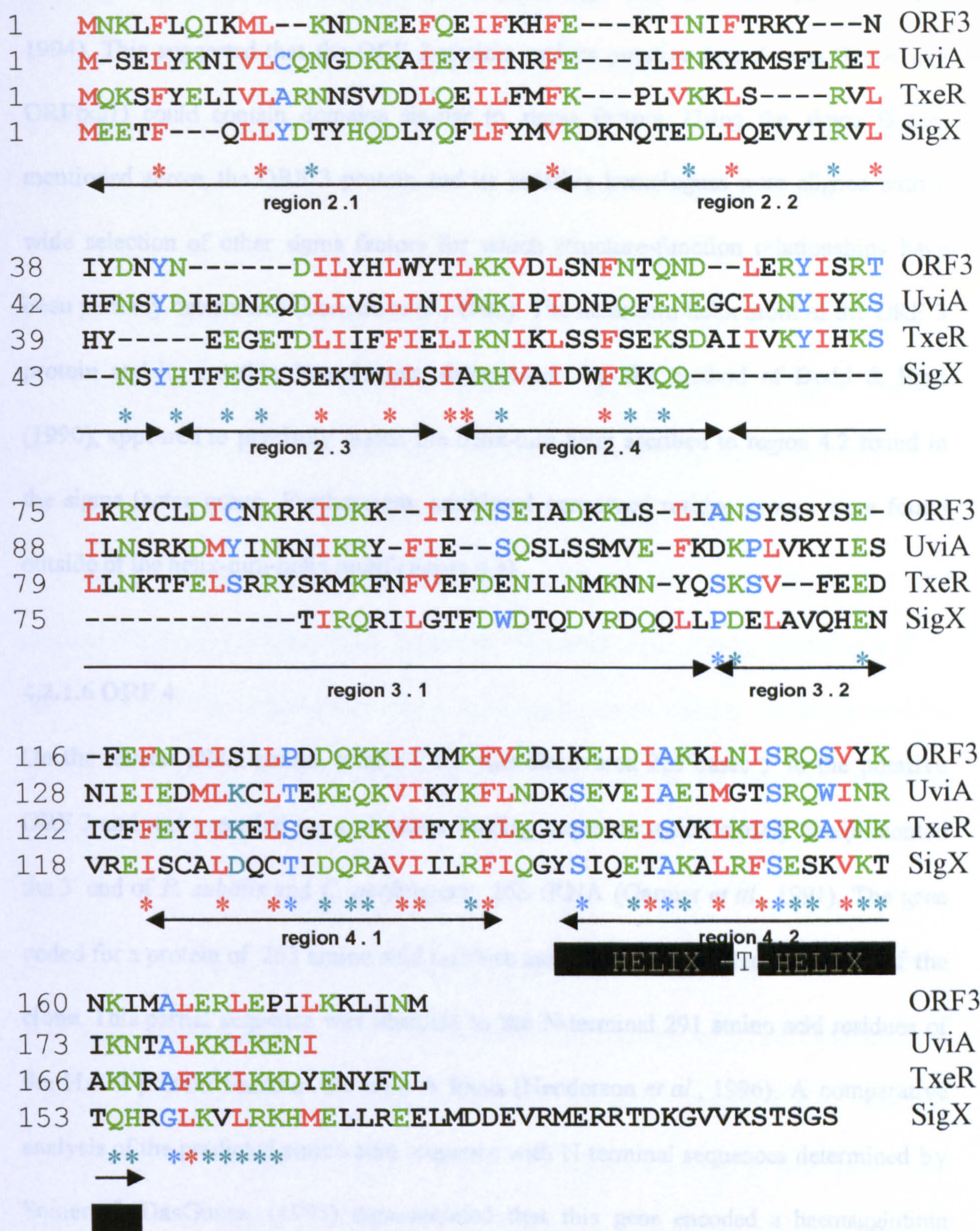


Figure 4.3 Alignment of ORF 3 polypeptide homologues with SigX

Alignment of the ORF 3 polypeptide, *UviA* and *TxeR* with *SigX* was performed using the DNASTAR software program Megalign by the Clustal method (Higgins & Sharp, 1988). Amino acids were grouped on a 'structural' basis where green implies external residues (DEHKNQR), red implies internal residues (FILMV) and blue implies ambivalent residues (ACGPSTWY). Residue groups are only shown in colour when the majority group is present. Asterices denote residue groups occurring in the σ^{70} family and the ORF 3 protein, *UviA* and *TxeR*, whilst arrows indicate functional domains which are conserved in the the σ^{70} related group with specific reference to *SigX*, residues 1-117 (Lonetto *et al.*, 1994) and residues 118-195, (Lonetto *et al.*, 1992).

(13.9 %) from *B. subtilis* (Yudkin *et al.*, 1987) and SigX from *B. subtilis* (Lonetto *et al.*, 1994). This suggested that the ORF 3 protein and its putative homologues (UviA and ORFtxe1) could contain domains similar to sigma factors. Using the sigma factors mentioned above, the ORF 3 protein and its putative homologues were aligned with a wide selection of other sigma factors for which structure-function relationships have been partially determined (Lonetto *et al.*, 1992). The helix-turn-helix motif of the ORF 3 protein and its putative homologues, determined by the method of Dodd & Egan (1990), appeared to precisely match the helix-turn-helix ascribed to region 4.2 found in the sigma factor group. Furthermore, additional conserved residue groups were found outside of the helix-turn-helix motif (figure 4.3).

4.2.1.6 ORF 4

On the reverse DNA strand, a third ORF was discovered 225 bases 5' to the putative ORF 3 and was preceded by a ribosome binding sequence which closely complimented the 3' end of *B. subtilis* and *C. perfringens* 16S rRNA (Garnier *et al.*, 1991). The gene coded for a protein of 263 amino acid residues and appeared truncated by the end of the clone. This partial sequence was identical to the N-terminal 291 amino acid residues of the HA34 protein found at the type A locus (Henderson *et al.*, 1996). A comparative analysis of the predicted amino acid sequence with N-terminal sequences determined by Somers & DasGupta, (1991) demonstrated that this gene encoded a haemagglutinin component identical to the type A neurotoxin complex described as the 35 kDa haemagglutinin protein (Somers & DasGupta, 1991) with 10 of 10 N-terminal amino acid residues matching. Comparison of HA34/B° with a 33 kDa haemagglutinin cloned and sequenced from *C. botulinum* type C (Tsuzuki *et al.*, 1990) demonstrated that the

gene products shared 33.0 % amino acid identity and were of a similar length (figure 4.4). In addition, the type A and type C HA33/4 genes were found in a similar orientation to the putative NTNHI and neurotoxin genes. They were thus deemed to be homologous. Comparison of HA34/B° with the current Swiss-Prot database (release 100) identified a region of 32.4 % similarity over a 93 amino acid overlap with the HA17/C gene (Figure 4.4) in the type C operon (Fujinaga *et al.*, 1994). A marginally lower percentage similarity was found with HA17/B° (section 4.2.1.8), a possible homologue of HA17/C. The conserved residues matching HA34/B° were essentially identical in both the type B and type C HA17 proteins (figure 4.4).

A second protein also showed a low percentage similarity with HA34/B°, the 100 kDa mosquitocidal toxin from *Bacillus sphaericus* (Thanabalu *et al.*, 1992), exhibiting a 28 % identity over a 245 amino acid overlap. Intriguingly, many residues which appeared to correlate between HA34/B° and HA17 found in type B and C, also were found to align with the mosquitocidal toxin (figure 4.4).

4.2.1.7 Origin of the neurotoxin complex sequences 5' to BoNT/A

A total of 7.996 kb of DNA were characterised in the region 5' to the start codon of the type A neurotoxin gene (Henderson *et al.*, 1996). The A+T content of the DNA was 75.1 % suggesting a clostridial origin. Within this sequence a total of five ORFs were identified by the similarity of their encoded proteins to other published proteins and N-terminal sequences, which may encode at least seven separate polypeptides. The putative genes appeared to be organised in three separate operons (figure 4.1). When analysing this data, a discrepancy was found between the theoretical position and actual position of *Xba*I and *Bgl*II restriction sites which were predicted by the Southern

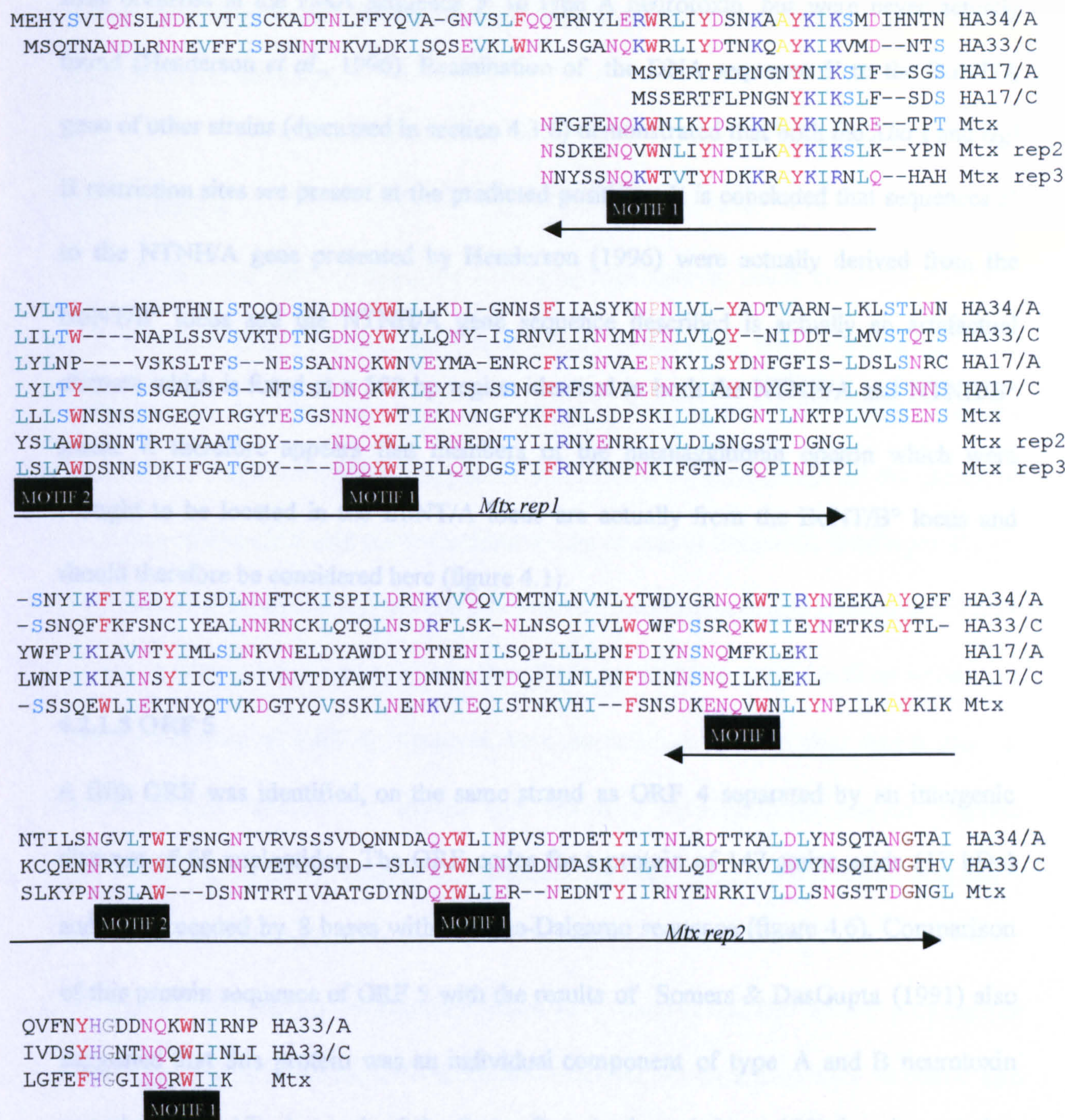


Figure 4.4 Comparative alignments of the HA34 haemagglutinin component of *C. botulinum* NCTC 2916.

Proteins which were similar to the 34 kDa haemagglutinin from *C. botulinum* NCTC 2916 were found using the DNASTAR software programs to scan the Swiss-Prot database (release 100). Homologues of approximately the same molecular weight were detected in other botulin neurotoxin regulons including type A NCTC 7272 (98.6 % identity) nonproteolytic type B (89.7% identity), (East *et al.*, 1994) and types C and D (identical proteins with 33 % identity), (Tsuzuki *et al.*, 1990; Ohyama *et al.*, 1995). A lower degree of similarity was found between type C HA17 protein (34 % identity over a 95 amino acid overlap)(Fujinaga *et al.*, 1994) and to a slightly lesser degree with the type A HA17 protein (this study). The 100 kDa mosquitocidal protein (Mtx) from *B. sphaericus* SSII-1 also displayed a significant degree of similarity (29.79 % identity over a 245 amino acid overlap). Alignment of the type A 34 kDa haemagglutinin (HA34/A) and type C 33 kDa haemagglutinin (HA33/C) with the type A and type C 17 kDa haemagglutinins (HA17/A and HA17/C) and partial Mtx coding sequence demonstrates a number of similar motifs and conserved residues which appear to occur in all 5 proteins. The sequences described as motif 1 and motif 2 appear to occur in multiple copies in the Mtx and HA33-34 kDa proteins and form part of a consensus motif occurring in the multiple internal repeat sequences (Mtx rep 1, 2 and 3) of Mtx (Thanabalu *et al.*, 1992). Amino acid groupings are (E, D, N, Q), (I, L, V), (W, Y, F), (S and T), (R and K), (A), (C), (P), (H), (M) and (G). Amino acids were grouped when the majority was at least three.

blotting data of Thompson *et al* (1990). The analysis of this data suggested that these two sites occurred in the DNA sequence 5' to type A neurotoxin, but were never actually found (Henderson *et al.*, 1996). Examination of the DNA sequence 5' to the BoNT/A gene of other strains (discussed in section 4.3.6) demonstrated that both the *Xba* I and *Bgl* II restriction sites are present at the predicted positions. It is concluded that sequences 5' to the NTNH/A gene presented by Henderson (1996) were actually derived from the BoNT/B^o locus and the NTNH/A gene sequence described is actually an artefactual chimera which is fused at a 500 bp region identical in both the NTNH/A and NTNH/B^o genes. It therefore appears that members of the haemagglutinin operon which were thought to be located in the BoNT/A locus are actually from the BoNT/B^o locus and should therefore be considered here (figure 4.1).

4.2.1.8 ORF 5

A fifth ORF was identified, on the same strand as ORF 4 separated by an intergenic distance of 68 nucleotides. The ORF codes for a protein of 147 amino acids (17 kDa) and was preceded by 8 bases with a Shine-Dalgarno sequence (figure 4.6). Comparison of this protein sequence of ORF 5 with the results of Somers & DasGupta (1991) also suggested that this protein was an individual component of type A and B neurotoxin complexes (HA17). A search of the Swiss-Prot database (release 100) found a putative homologue of HA17 in the type C botulinal neurotoxin complex locus (Fujinaga *et al.*, 1994). The type C HA17 was of identical length and shared 63 % amino acid identity with HA17/B^o. HA34 and HA17 also shared a similar organisation to their counterparts at the type C locus (Fujinaga *et al.*, 1994) and were separated by virtually identical intergenic gaps. However, the 17 kDa proteins of the BoNT/B^o and type C progenitor

neurotoxin complexes contained a higher degree of amino acid sequence identity in comparison to the larger 34 kDa and 33 kDa proteins of the BoNT/B^o and type C progenitor neurotoxin loci.

4.2.1.9 ORF 6

A sixth ORF was discovered closely associated with ORF 5, which coded for a protein of 625 amino acids (71.14 kDa). The initiation codon was separated from a Shine-Dalgarno-like sequence by seven nucleotides, which in turn was adjacent to the termination codon of ORF 5 (Appendix III). An initial comparison of the predicted sequence of this ORF with the N-terminal sequence data of Somers & DasGupta (1991) indicated that the remaining two peptide sequences corresponding to proteins of 57 kDa and 21.5 kDa in the neurotoxin complex. Both appeared to originate from within the coding sequence of ORF 6. However, both occurred at internal sites which did not correlate with translational signals and therefore were possibly due to multiple proteolytic processing. Indeed, the predicted cleavage sites both occurred between a lysine and valine residue.

A search of the Swiss-Prot database (issue 100) revealed that the ORF 6 polypeptide had a homologue in the type C neurotoxin complex which shared 67.7 % amino acid identity. The corresponding ORF found in the type C locus (Fujinaga *et al.*, 1994) was orientated in an identical manner to ORF 6 and was separated from the HA17 gene by 16 bp, an identical intergenic distance was found between ORF 5 and ORF 6. Both proteins also appeared to result from post translational modification where up to 15 hydrophilic amino acids were removed at the amino terminal ends, whilst a second cleavage created a 21.5 kDa and 53 kDa moiety. ORF 6 was also found to share 20 %

amino acid identity with the *C. perfringens* type A enterotoxin (CPE) (Czeczulin *et al.*, 1993) over a 202 amino acid overlap (figure 4.5).

4.2.2 Repetitive sequences

Haemagglutinins and other lectins are commonly found to contain repetitive domains (Romeo *et al.*, 1986). Repeating motifs were searched for using the 'Dot-Matrix' algorithm method (DNASTAR inc.). Self alignment of the 3 genes occurring in the putative haemagglutinin operon exposed a number of possible repetitive sequences. The 21.5 kDa component of the HA70 protein appeared to share sequence identity with the contiguous 53 kDa component (figure 4.5). Elements of this similarity were also shared with CPE. This relationship was previously shown to occur in the *C. botulinum* type C HA70 protein (Fujinaga *et al.*, 1994).

A repeating motif was discovered in the coding sequence of the HA 34 protein (figure 4.4). This occurred 4 times, and in two cases was preceded by a similar repeat (figure 4.4), and intriguingly, the double repeat also occurred in the HA 17 protein. Alignment of HA 34 and HA17 indicate that these two sequences could share an evolutionary relationship (figure 4.4). Motifs similar to these also occur in the 3 regions of internal homology found in the *B. sphaericus* mosquitocidal toxin (Thanabalu *et al.*, 1992).

4.2.3 Transcripts

Northern blotting data generated by Dr I. Henderson (Henderson *et al.*, 1996) demonstrated that radio-labeled DNA probes derived from either the HA70, HA17 or HA34 genes detected a single 3.2 kb transcript. Transcription of the putative haemagglutinin genes therefore, appeared to be linked

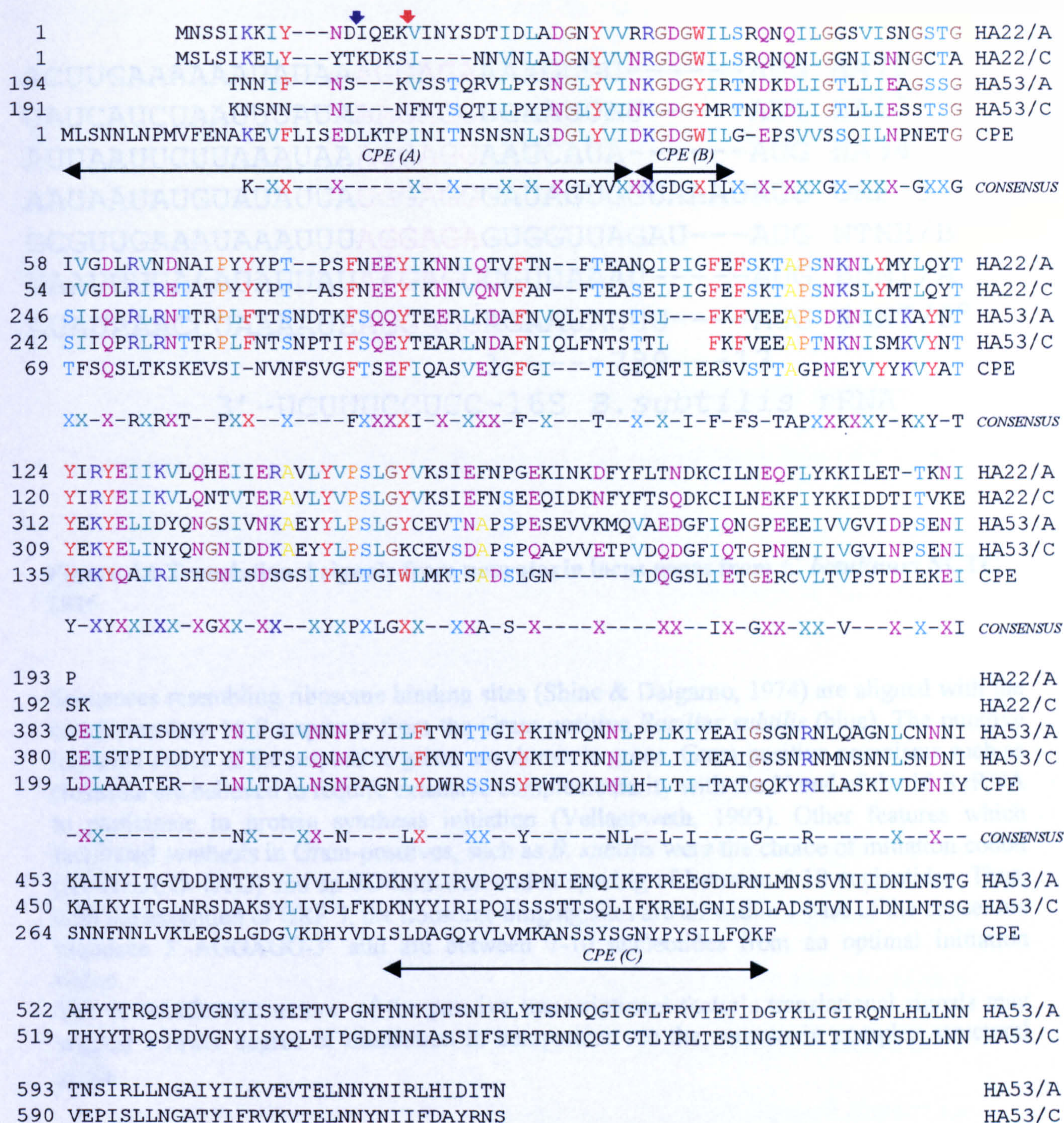


Figure 4.5 Alignment of HA70 derivatives with the *C. perfringens* CPE enterotoxin

The HA70 'Haemagglutinin' protein sequence from *C. botulinum* NCTC 2916 shares a 67.7 % amino acid identity with a homologue in the type C neurotoxin locus (Fujinaga *et al.*, 1994). Evidence suggests that both proteins undergo proteolytic cleavage into 53 kDa and 22 kDa entities (Fujinaga *et al.*, 1994; Inoue *et al.*, 1996). The cleavage product of type A may be further processed (blue arrow indicates HA53 cleavage and red arrow indicates HA22 cleavage (Inoue *et al.*, 1996). Alignment of the 53 kDa and 22 kDa moieties from both type A and type C proteins demonstrates that a degree of sequence conservation exists. At the N terminus of all peptides exists a conserved motif surrounding the tri-peptide sequence GDG; this region (CPE(B)) is considered essential for the toxic action of cpe whilst proteolytic removal of region CPE A enhances activity (Kokai-Kun & McClane, 1997). The C-terminal region of CPE (CPE(C)) encodes the enteric receptor domain (Kokai-Kun & McClane, 1997) and does not appear to resemble the C-terminus of the botulin 53 kDa fragments. Amino acid groupings are (E, D, N, Q), (I, L, V), (W, Y, F), (S and T), (R and K), (A), (C), (P), (H), (M) and (G). Amino acids were grouped when the majority was at least three.


```

ACUUGAAAAAAUAUAAAGGAGAAAAGUGU-----AUG HA70
UAUCAUCUAAUUUAUAAGAGGUGAAUUAU-----AUG HA17
AUUAAUUCUUAUAAUAAAGGAGGAAUCAUA-----AUG HA34
AAUAAUAUGUAUAUUAUGGAGGGAUAGUGGUAAAU AUG ORF 3
GCGUUGAAAUAAAUUUAGGAGAGUGGUUAGAU---AUG NTNH/B
UAAUUUUUAAAUAUUAUAAGAGGUGUUAUAAU-----AUG BoNT/A
AUAUAAACUUAUAAAUAAAGGAGGAGAAUAUUU-----AUG BoNT/B°
                                1-----789---13
                                3'-UCUUUCCUCC-16S B. subtilis rRNA

```

Figure 4.6 Translational signals from neurotoxin locus genes from *C. botulinum* NCTC 2916

Sequences resembling ribosome binding sites (Shine & Dalgarno, 1974) are aligned with the complementary 16 S sequence from the Gram-positive *Bacillus subtilis* (blue). The putative initiation codon of the respective genes are shown in green. Gram-positive organisms such as clostridia are believed to require extensive complementarity with the 3' end of the 16 S rRNA to participate in protein synthesis initiation (Vellanoweth, 1993). Other features which facilitated synthesis in Gram-positives, such as *B. subtilis* were the choice of initiation codon (AUG>UUG>GTG) and an SD-initiation codon spacing of between 6-10 nucleotides. Thus, with the exception of ORF 3, the ribosome binding sites are all within 1 base of the consensus sequence 5'-AGGAGG-3' and are between 7-10 nucleotides from an optimal initiation codon.

The non-conformist nature of the putative transcriptional factor's translational signals may suggest a lower degree of translation in comparison to the neurotoxin complex structural genes.

ORF 3 appears to form a monocistronic operon. The gene is divergently transcribed from the putative haemagglutinin operon and is 5' to the proposed NTNH-neurotoxin operon promoter. Therefore, it is unlikely to be under the direction of either of these promoters. Although a transcript of ORF 3 was not detected by Northern blotting (Dr I. Henderson pers. comm.) a cDNA product was detected, which mapped 64 bases 5' to the ORF 3 initiation codon and from which -35 (TGGATA) and -10 (TATTAA) boxes could be tentatively ascribed (Henderson *et al.*, 1996).

ORF 3 is closely followed by the putative promoter regions regulating the expression of the NTNH gene (Appendix III). However, transcriptional read through from ORF 3 into the following operon does not seem to contribute significantly, as Northern blotting of total RNA from *C. botulinum* NCTC 2916, using ORF 3 derived DNA as a probe, did not detect the larger 7.5 kb NTNH-neurotoxin transcript.

4.2.4 Transcriptional signals

The divergent nature of the genes encoding the neurotoxin complex components implied that a minimum of two promoters must be present. Comparative alignment of the intergenic sequences preceding the NTNH gene and putative HA 34 gene identified a common sequence of 23 bp which was followed by a second conserved element of 10 bp separated by 7 bp (see figure 4.8 and Appendix III). The larger sequence contained a motif 'TTTACA' which is virtually identical to the -35 promoter element recognised by the major form of RNA polymerase (Young *et al.*, 1989). Subsequent primer extension analysis by Dr I. Henderson located mRNA initiation sites 3' to the putative promoter sequences in front of both the NTNH and putative HA 34 kDa genes.

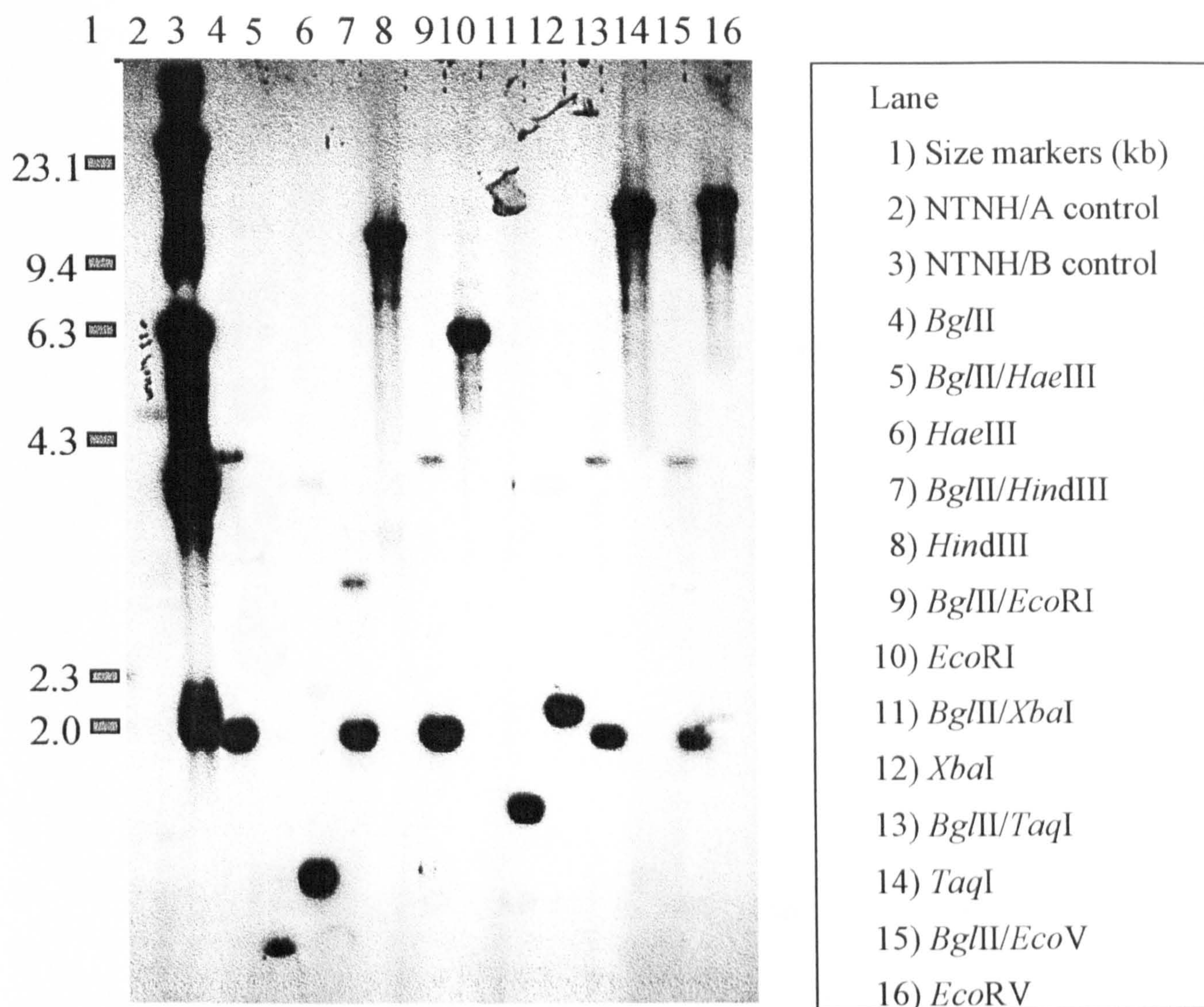


Figure 4.7a Generating a genomic restriction map of the type B neurotoxin locus in *C. botulinum* NCTC 2916

Genomic digests of chromosomal DNA isolated from *C. botulinum* NCTC 2916 were separated in an agarose gel and transferred to a Nylon membrane. The membrane was probed with probe 3, a radiolabeled *Bgl*II-*Stu*I DNA restriction fragment isolated from a PCR clone of the type B NTNH gene from the same strain (see figure 4.1). The probe was hybridised at 56 °C for 16 h in 'GOOP' hybridisation solution. Minor bands typically observed near 4.4 kb probably result from weak interaction with the type A NTNH gene.

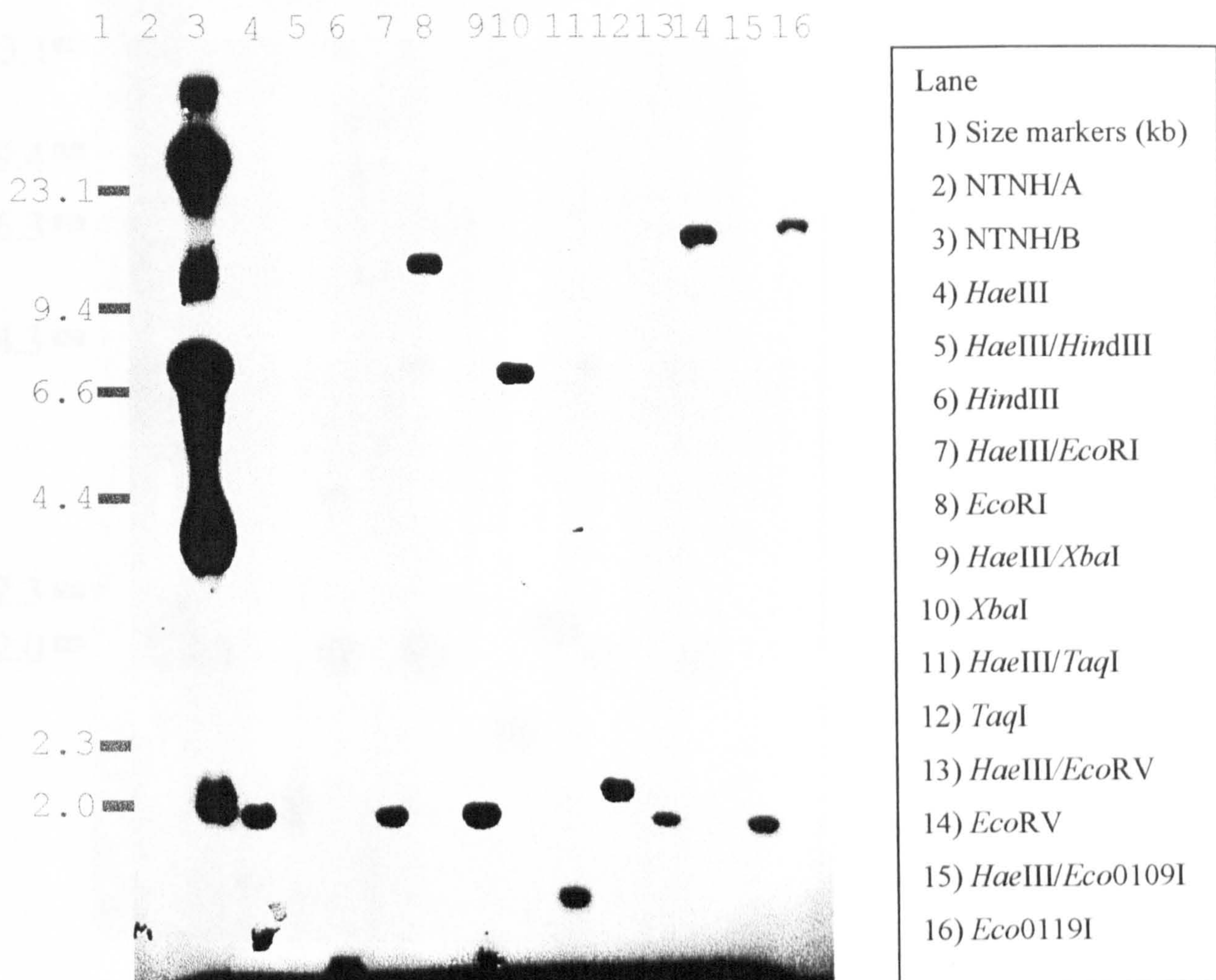


Figure 4.7b Generating a genomic restriction map of the type B neurotoxin locus in *C. botulinum* NCTC 2916

Genomic digests of chromosomal DNA isolated from *C. botulinum* NCTC 2916 were separated in an agarose gel and transferred to a Nylon membrane. The membrane was probed with probe 3, a radiolabeled *Bgl*II-*Stu*I DNA restriction fragment isolated from a PCR clone of the type B NTNH gene from the same strain (see figure 4.1). The probe was hybridised at 56 °C for 16 h in 'GOOP' hybridisation solution.

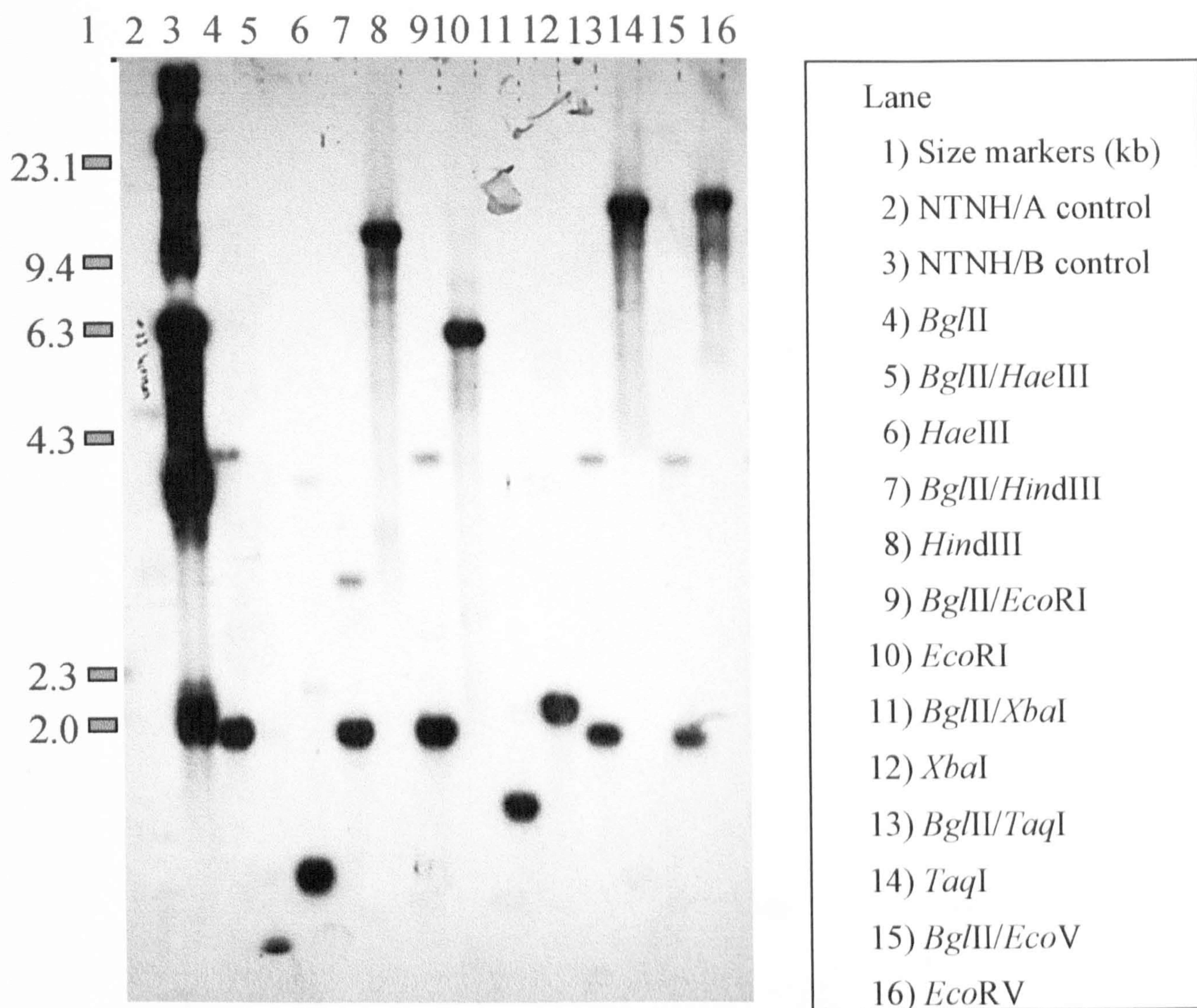


Figure 4.7a Generating a genomic restriction map of the type B neurotoxin locus in *C. botulinum* NCTC 2916

Genomic digests of chromosomal DNA isolated from *C. botulinum* NCTC 2916 were separated in an agarose gel and transferred to a Nylon membrane. The membrane was probed with probe 3, a radiolabeled *Bgl*II-*Stu*I DNA restriction fragment isolated from a PCR clone of the type B NTNH gene from the same strain (see figure 4.1). Positive NTNH type A and type B controls were pCBA4 and pSB17 respectively (see p112, figure 4.1). The probe was hybridised at 56 °C for 16 h in 'GOOP' hybridisation solution. Minor bands typically observed near 4.4 kb probably result from weak interaction with the type A NTNH gene.

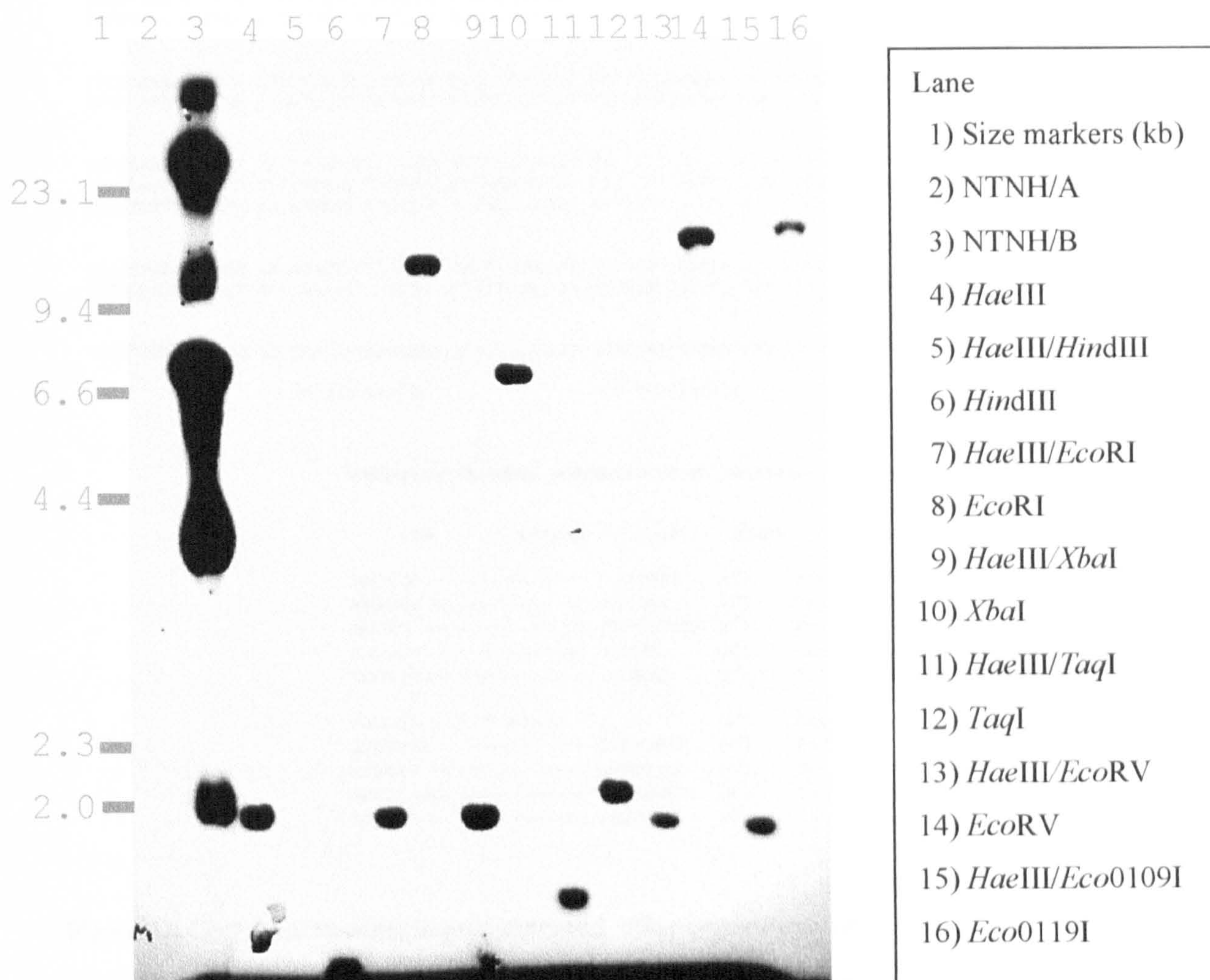


Figure 4.7b Generating a genomic restriction map of the type B neurotoxin locus in *C. botulinum* NCTC 2916

Genomic digests of chromosomal DNA isolated from *C. botulinum* NCTC 2916 were separated in an agarose gel and transferred to a Nylon membrane. The membrane was probed with probe 3, a radiolabeled *Bgl*II-*Stu*I DNA restriction fragment isolated from a PCR clone of the type B NTNH gene from the same strain (see figure 4.1). Positive NTNH type A and type B controls were pCBA4 and pSB17 respectively (see p112, figure 4.1). The probe was hybridised at 56 °C for 16 h in 'GOOP' hybridisation solution.

	-35		-10				
TTATATTTTAAAAA	ATTTTAGGTTTACAAAA	AATAGTGTGGCT	ATGTTATATATAA	ATG-	84nt-AGGAGA-	6nt-ATG-NTNH/A	Align.1
ATTTGATATATTTAA	ATTTTAGGTTTACAAAA	AATAACTTGATT	ATGTTATATGTTA	TAT-	21nt-AGGAGG-	7nt-ATG-HA34/A	"
TTTATATATTTTAT	ATTTTAGGTTTACAAAA	GACGATGAAGAA	ATGTTATATATAA	GTG-	82nt-AGGAGA-	10nt-ATG-NTNH/C1	"
AATAGATATATTATA	ATTTTAGGTTTACAAAA	TTTGGCTTAAAT	ATGTTATATGTAA	GTG-	20nt-GGGAGG-	6nt-ATG-HA33/C	"
TTTTTAAAAATTTAA	ATTTTCAGTTTACAAAA	AATAACCTGATT	ATGTTATATGTAA	TTG-	110nt-AGGAGA-	8nt-ATG-TeNT	"
TATATCAAATTTTTA	TTAGTATGTTTACATAA	TTGATTATGGAT	ATTCGTAAAAAT	GGC-	62nt-TGGAGG-	13nt-ATG-ORF 3	Align.2
TAAATAATATTATAG	AGAATTAGTTTACATAT	TATTACATGTAA	GTTATAATAATGT	AAA-	5nt-AGGGGA-	7nt-ATG-ORF22/C	"
AAATAAAAAAATTAT	AAATTTAGTTTACAAAA	TTGAAGTCAAAT	TACTTTTATATT	ATG-	74nt-GAGAGG-	8nt-ATG-Bcn5 P1	Align.3
TATCGGCTTTTAA	TTTTTAGGTTTACATTT	TTAAACTAAA	CTCTTTTATTTA	TTA-	156nt-GAGAGG-	8nt-ATG-Bcn5 P3	"
TAGATTTTATAAAAA	AATAGATATTTACAAAA	TAGACTAAAAAA	AGCTTTTATATA	N ₁₁ T-	152nt-AGGGGT-	9nt-ATG- <i>uviA</i> P5	"
ACATTAAAAAATAT	AAGATATGTTTACAAAT	TACTATCAGACA	ATCTCCTTATCTA	N ₄ A-	222nt-AGGAGG-	5nt-ATG-ToxA	Align.4
TTTATATTTTATATA	GAACAAGTTTACATAT	TTATTTCAGAAA	AAGTCTTTATTCA	ATC-	151nt-AGGAGA-	7nt-ATG-ToxB	"
TATTTATTGACTAAA	TTATAAGTTTACATAA	TTATTTAATAAT	TATGTAATTGTTA	CTT-	333nt-AGAGGG-	7nt-ATG-TxeR	Align.5

-35 similarity

-10 similarity

CONSENSUS PROMOTER SEQUENCES OF *B. subtilis* σ FACTORS

-35	SPACER	-10	SIGMA	FUNCTION in <i>B. subtilis</i>
TTGACA---	-----	--TATAAT	(σ^A)	House keeping/early sporulation.
RGGXTTRA-	-----	-GGGTAT	(σ^B)	General stress response.
AAATC----	-----	T AXTGYTTZTA	(σ^C)	Post exponential gene expression.
TAAA-----	-----	GC CGATAT	(σ^D)	Chemotaxis/autolysis/flagellar exp.
RWAGGAXXT	-----	--HGAAT	(σ^H)	Post exponential gene expression, competence and early sporulation.
TGGCAC---	--TTGCANNN		(σ^L)	Degradative enzyme gene expression.
ZHATAXX--	-----	CATACAHT	(σ^E)	Early mother cell gene expression.
GCATR----	-----	G GHRARHTX	(σ^F)	Early forespore gene expression.
GHATR----	-----	--CATXHTA	(σ^G)	Late forespore gene expression.
AC-----	-----	CA TANNNTA	(σ^K)	Late mother cell gene expression.

Figure 4.8 Clostridial neurotoxin promoters and other cognate promoters

Alignment 1 consists of mapped clostridial neurotoxin operon promoters (adapted from Henderson *et al.*, 1996) which consist of the promoter regions directing expression of the haemagglutinin operon in type A; HA33/A and type C; HA34/C, the NTNH-neurotoxin operon from type A; NTNH/A and type C; NTNH/C and for the *C. tetani* neurotoxin; TeNT.

Alignment 3 and 4 show mapped promoters from the *C. perfringens* BCN5 regulon (Garnier & Cole, 1988) and *C. difficile* toxigenic element (Hammond *et al.*, 1997), respectively, where promoters P1 and P3 direct expression of the bacteriocin 5, P5 directs expression for the putative transcriptional factor *uviA*, and *tox4* and *tox8* direct the expression of the *C. difficile* type A and B enterotoxins and cytotoxins, respectively.

Alignment 2 shows unmapped sequences preceding ORF3 and its homologue in the type C neurotoxin locus ORF 22 (Hauser *et al.*, 1994), whilst alignment 5 shows a similar unmapped sequence in front of the putative positive transcriptional regulator TxeR of the *C. difficile* toxigenic element (Hammond *et al.*, 1997; Moncrief *et al.*, 1997).

The comparisons appear to show that an extended -35 type sequence exists in the clostridial neurotoxin operons surrounding a core sequence TTTACA, which is similar to the -35 σ^A consensus promoter element. Features of this extended promoter (RRRTTTACAAXX) appear to be present in the BCN5 and *C. difficile* cytotoxin regulons (alignments 3, 4 and 5) which also possess putative homologues of ORF3. However the highly conserved -10 box found in the clostridial neurotoxin regulons may be replaced with alternative 'specific' -10 analogues in the other operons. It is interesting to note that sequences resembling the -35 box occur in front of ORF 3, ORF22 (alignment 2), *uviA* (alignment 3; P5) and TxeR (alignment 5). These could participate in some type of autoregulation.

Other characterised sigma factor consensus sequences are shown below (adapted from Haldenwang, 1995) Ambiguous bases are coded; H: A or C, N: A, G, C or T, R: A or G, W: A, G or C, X: A or T, Y: C or T, Z: T or G. Transcriptional start sites are shown in magenta.

4.3 Discussion

4.3.1 Haemagglutinin genes

ORFs 4, 5 and 6 appear to form an operon encoding haemagglutinin proteins. A similar operon has been characterised in the type C neurotoxin locus (Fujinaga *et al.*, 1994; Hauser *et al.*, 1994) and by analogy has also been characterised to varying degrees in representatives of serotypes A and B (East *et al.*, 1994), B (Yang *et al.*, 1996), (Bhandari *et al.*, 1997), A(B^o) (Hutson *et al.*, 1996), D (Ohyama *et al.*, 1995) and G (Bhandari *et al.*, 1997). Serotypes E, F and some sub-types of serotype A described as A₂ do not appear to possess a haemagglutinin operon closely associated with the respective neurotoxin locus (East *et al.*, 1996). The existence of haemagglutinins in *C. botulinum* serotypes and their association with the neurotoxin complex is a well established phenomenon (Lowenthal & Lamanna, 1952; Oguma *et al.*, 1976). Consequently, the genetic alliance of genes encoding the neurotoxin and haemagglutinins ties in well with these earlier observations. However, haemagglutination has not been directly established for each individual component. The HA 34 gene would appear to encode a haemagglutinin. Tsuzuki was able to demonstrate that deletions in the HA 33/C gene cloned in *E. coli* eliminated the haemagglutinating activity which had been demonstrated for the intact gene (Tsuzuki *et al.*, 1990). Haemagglutinin activity was attributed to the type G neurotoxin complex (Nukina *et al.*, 1991) which does not appear to encode a homologue of the HA 33 gene (Bhandari *et al.*, 1997). Thus, on the basis of a 'simple' comparison, it could be deduced that the HA 33 homologues were not entirely responsible for all haemagglutinating activities of the neurotoxin complex. In the report of Somers & DasGupta (1991), individual elements of the type A or B neurotoxin complex corresponding to the 53 kDa, 35 kDa, 21.5 kDa and 17 kDa

components could not demonstrate haemagglutination of RBC. This was attributed either to the presence of SDS, or to the fact that heterologous haemagglutinins were required containing multiple ligand binding sites to effect haemagglutination. However, with the observations of Tsuzuki (1990) in mind, this could not be a logical explanation of why the HA 34 did not function.

Close inspection of the HA 34 coding sequence revealed a number of repeating features, characterised by a tryptophan residue preceded by residues with polar side chains and followed by another hydrophobic residue (figure 4.4). Two of these repeats appear to be preceded by a second motif also containing a tryptophan residue (figure 4.4). These motifs form a significant proportion of residues common between the HA 34 and HA 17 homologues from other botulinal neurotoxin complexes and the 3 repeating domains of the mosquitocidal toxin from *B. sphaericus* (Thanabalu *et al.*, 1992). As these motifs appear to occur in multiples, it is tempting to speculate that they may contribute to the binding capacity of the clostridial haemagglutinins and mosquitocidal toxin. Furthermore, the sequence similarity detected between the HA 34 and HA 17 proteins indicate that they may have common ancestry, where the HA 34 gene might have arisen by a duplication of the HA 17 gene. Domain duplication has been implicated in the formation of other haemagglutinins (Romeo *et al.*, 1986). Analysis of the predicted secondary structures of the haemagglutinin proteins and the mosquitocidal toxin demonstrate that they are all mainly composed of β -pleated sheets punctuated by a large number of turns and coils. The structures are predicted to contain flexible regions along their entire length. A similar secondary structure has been described for the M-agglutinin of *E. coli* and certain lectins (Rhen *et al.*, 1986).

Comparison of the HA 70 coding sequence failed to detect motifs found in the HA 34 protein. However, Fujinaga first reported a primary sequence similarity between the proteolytic cleavage products of the HA 70 protein of the type C neurotoxin complex and the *C. perfringens* type A enterotoxin (CPE) (Fujinaga *et al.*, 1994). A similar relationship is also evident from this study. Several interesting parallels exist between these proteins. Firstly, post-translational modification of HA 70 produces HA 21.5 and HA 53 proteins. The HA 21.5 kDa protein is then further processed from the N-terminus in *C. botulinum* type A (Fujita *et al.*, 1995) and type C (Fujinaga *et al.*, 1994). The CPE enterotoxin can also be modified by peptidases endogenous to the human digestive system, removing up to 44 N-terminal amino acids and concomitantly increasing cytotoxicity (Kokai-Kun & McClane, 1997). They are all predicted to comprise β -pleated sheets, interspersed with turns and coils, and are flexible. In addition, the CPE enterotoxin, HA 21.5 and HA 53 proteins at some point all specifically interact with other proteins (Kokai-Kun & McClane, 1997).

Details of the structure-function relationship of CPE enterotoxin are emerging. The 30 C-terminal amino acids are thought to determine initial receptor binding to a 50 kDa protein, the initial intestinal target of the enterotoxin (Kokai-Kun & McClane, 1997). At the N-terminal end, residues 45-53 are essential for cytotoxicity (Kokai-Kun & McClane, 1997). The N-terminal 45-53 residues of the enterotoxin also form part of a well conserved motif common to the type A and C HA 53 and HA 21.5 proteins (figure 4.5). At the center of the motif, a RGD sequence present in both HA 21.5 proteins (but replaced by KGD in the HA 53 and CPE proteins) was suggested as a cell attachment site (Fujinaga *et al.*, 1994).

Pairwise alignment of the C-terminal end of the CPE enterotoxin with the C-terminal end of the HA 53 protein show negligible similarity. In conclusion, the similarity between CPE and the HA 53 and HA 21.5 appears to indicate a distant but common ancestry, however the extrapolation of functions to the botulinum proteins would be naively ambitious.

The HA 70 gene may have been formed by a duplication event. However, although the exact function of the resultant proteins from this gene appears obscure, but logically would in some way facilitate the delivery of the neurotoxin to the site of intestinal absorption. Whether it also has a lectin-like activity or binds to other (protein?) elements of the intestinal membrane, as well as the neurotoxin complex, remains a mystery.

4.3.2 ORF 3 ~ A putative transcriptional factor?

4.3.2.1 Similarity to the σ^{70} family of proteins?

Sigma factors are unified by their ability to confer promoter-specific transcription initiation on RNA polymerase. The σ^{70} related proteins are composed of specific domains for which some structure-function relationships have been determined (Lonetto *et al.*, 1992). However, the degree to which these domains are present and the extent of sequence conservation varies. This variation is thought to reflect the role served in their respective hosts (Lonetto *et al.*, 1992). Among the domains which have been described, region 2.1 is thought to mediate core polymerase binding, region 2.3 is involved in the interaction with the -10 promoter sequence and region 4.2 forms a helix-turn-helix motif which specifically recognises the -35 box of a promoter (Lonetto *et al.*, 1992).

The most convincing similarity between the ORF 3 polypeptide, UviA, the ORFtxe1 polypeptide and a compilation of sigma's presented by Lonetto (1992) is found in regions 4.1 and 4.2. Not only are a number of highly conserved residues present, but other features resemble sigma factors, including a basic charge consistent with DNA binding, a predominately α -helical secondary structure and a C-terminal helix-turn-helix motif.

Beyond region 4, the extrapolation of domain motifs to recognised regions of the sigma factors is obscure, so the contribution of approximately the first 140 amino acids of the ORF 3 protein is difficult to predict.

4.3.2.2 The helix-turn-helix motif

This motif was predicted using the method of Dodd & Egan (1990), (figure 4.2). Promoter specificity has been demonstrated to be particularly influenced by amino acid residues at positions +3, +4, and +7 relative to the center of the H-T-H motif (Lonetto *et al.*, 1992) and it is interesting to note that in the ORF 3 protein, UviA and the ORFtxe1 protein, an identical motif, SRQ, is found at positions +2 to +4 in the H-T-H.

4.3.2.3 Parallels with the UviA and ORFtxe1 loci

UviA and the ORFtxe1 protein represent the closest homologues of the ORF 3 protein outside of the putative clostridial neurotoxin 21 kDa transcriptional factor group. As the promoters of this group seem to be extremely well conserved among the clostridial neurotoxin operons, the promoter regions of genes associated with *uviA* and ORFtxe1 were examined. Both regulons contained mapped transcriptional start sites with potential -35 elements which have considerable identity to the clostridial neurotoxin

operon promoters (Garnier & Cole, 1988; Hammond *et al.*, 1997). For *uviA*, the promoter occurs in front of the *bcn* gene encoding the bacteriocin BCN5, and, for ORFtxe1, the promoters are in front of the *C. difficile* cytotoxin A and cytotoxin B genes (*toxA* and *toxB* genes, respectively).

Extrapolation of mutagenesis experiments on the sigma factor group (Lonetto *et al.*, 1992) would suggest that the amino acid sequence SRQ found in the H-T-H motifs of UviA, ORFtxe1 and ORF 3 proteins could contribute towards this promoter specificity. Intriguingly, a very similar -35 element has been mapped in front of *uviA* and, on examination of the nucleotide sequence, could also be present in front of ORFtxe1, ORF 3 and its homologues from type A, B and C neurotoxin complex operons. This raises the possibility that these putative transcriptional factors may participate in some type of autoregulation. Recently, the product of ORFtxe1 (TxrR) was shown to positively regulate *C. difficile toxA* and *toxB* promoters in *E. coli* (Moncrief *et al.*, 1997).

4.3.2.4 A member of the FixJ family?

Several authors have grouped homologues of ORF 3 in the FixJ family on the basis that the ORF 3 botulinum homologues are similar to UviA which in turn has been grouped in the FixJ family (Lonetto *et al.*, 1992). At the present moment, this categorisation does not appear to be valid due to several observations. A report which separates the FixJ family from the σ factors documents the unique features found in the FixJ sequences (Kahn & Ditta, 1991). Features which were purported to define the FixJ response regulators were not generally observed in the ORF 3 polypeptide and UviA. In particular, a conserved glycine residue at position 156 of the FixJ family (Kahn & Ditta, 1991) was not present in the ORF 3 protein or UviA (although it is present in a number

of group 3 sigmas; Lonetto *et al.*, 1992), a conserved threonine was not found at the equivalent position 172 (Kahn & Ditta, 1991), neither the ORF 3 protein or UviA were of sufficient size to contain the 191-200 region of FixJ members, did not possess an aromatic residue at 'FixJ position' 176 or a hydrophobic residue at 'FixJ position' 186 (Khan & Ditta, 1991). The inference that ORF 3 encoded a FixJ member would have insinuated that the N-terminal portion was specifically an acceptor module, typically modulated by a kinase sensor (Khan & Ditta, 1991).

An interesting tangent originating from the FixJ question is how much of the ORF 3 protein is required for its putative transcriptional function? FixJ was bisected into FixJN₍₁₋₁₂₆₎ and FixJC₍₁₃₁₋₂₀₄₎. The 73 amino acid moiety FixJC was capable of directing promoter specific expression. Thus, the relatively limited section surrounding the helix-turn-helix motif must be capable of interacting with the core RNA polymerase (Khan & Ditta, 1991).

4.3.2.5 A member of the ECF σ family?

A number of transcriptional factors have been grouped on the basis that they exert transcriptional control on a regulon whose protein product(s) have an extracytoplasmic function (ECF) (Lonetto *et al.*, 1994). Although this name probably does not strictly apply to all of the members (Huang *et al.*, 1997), the concept of a disparate group of small sigma's effecting control on a contiguous regulon seems to be useful.

The ECF group is characterised at the structural level by conserved regions and deletions when compared to the larger sigma factors (Lonetto *et al.*, 1994). Alignment of SigX, a member of the ECF group, with UviA and the ORF 3 and ORFtxe1 proteins, using amino acid groupings devised either on a structural, chemical, charge or functional

basis, produced virtually identical alignments which revealed a number of conserved residue groups. This may indicate that regions equivalent to 2.1, 2.2, 2.3 and 2.4 present in SigX (Lonetto *et al.*, 1994) may exist in the ORF 3 protein sub-family. However, the inherently low sequence conservation between these types of transcriptional factors underlies the tenuous nature of these deductions (figure 4.3).

Region 2.1 of sigma factors is thought to participate in core polymerase binding. Here, high primary amino acid sequence identity is not expected as different sigma factors are thought to interact with varying affinities with the core RNA polymerase, and, therefore, have non-identical core binding sites (Gross *et al.*, 1992). Region 2.1 and 2.3 function in DNA strand melting (Lonetto *et al.*, 1992). Amino acid residue groups occur in the ORF 3 group which are highly conserved in the ECF family in these two regions. A similar situation occurs in the region equivalent to 2.4 in the ECF family, implicated in sequence specific contacts with the -10 promoter region. Again, a very high sequence conservation would also not be expected if different -10 promoter elements were recognised.

A region equivalent to 4.1 and 4.2 appeared to be well conserved with the ECF and other classes of sigma factors, but the ORF 3 sub-family did not extend in the C-terminal direction beyond the H-T-H motif, as seen in members of the ECF and σ^{70} family.

Significant differences between ECF, σ^{70} and the ORF 3 sub-family were found in region 3.1 and region 3.2. This is a region which has little sequence conservation among different classes of sigma factors and is tentatively implicated in core RNA polymerase interactions.

On the basis of a comparison between the ORF 3 protein and SigX, both are not predicted to contain a 1.2 domain. When 1.2 is present, it is thought to obstruct the sigma factor from binding to promoter sequences in the absence of the core RNA polymerase (Lonetto *et al.*, 1994, Dombroski *et al.*, 1992). Thus, the ORF 3 protein may bind to its promoter autonomously.

In conclusion, the ORF 3 protein, UviA and the TxeR could not be easily confined to an existing transcriptional sub-family and could form the basis of a novel group. Although H-T-H motifs are found in a variety of transcriptional regulators, including both transcriptional repressors and activators (Pabo & Sauer, 1992), the G-terminal location of the H-T-H and appropriately positioned conserved residues suggests that the ORF 3 protein, UviA and TxeR are comparable to group 3 of the σ^{70} family (Lonetto *et al.*, 1992). These proteins typically act as positive transcriptional factors of individual regulons. Interestingly, the encoded 21.6kDa product an ORF which is spatially associated with the *adc* operon of *C. acetobutylicum* appears to bear a low degree of similarity with this group (Gerischer & Durre, 1990).

4.3.3 Promoter structure

Primer extension studies by Dr I. Henderson were able to positively identify promoter structures corresponding to the divergently orientated operons coding for structural genes of the type A neurotoxin locus (Henderson *et al.*, 1996). The conserved features 'box 1' and 'box 2' appeared to represent -35 and -10 elements of a promoter (Hawley & McClure, 1983). At the core of the -35 element, a sequence (TTTACA) was noted that closely resembled the vegetative promoter (TTGACA) recognised by σ^{70} in *E. coli* and σ^A in *B. subtilis*. In front of either operon, box 1 was composed of a highly

conserved sequence of 17 nucleotides. Intriguingly, box 1 was almost perfectly conserved in front of both structural neurotoxin gene operons from *C. botulinum* type C (Fujinaga *et al.*, 1984), proteolytic and non-proteolytic type B *C. botulinum* strains (East *et al.*, 1994) and in front of the *C. tetani* neurotoxin gene (Eisel *et al.*, 1986). In all examples where box 1 was found, a second conserved sequence box 2, was separated by exactly 12 nucleotides. Box 2 was well represented with A and T residues typically found at -10 promoter sites consistent with a role in DNA strand melting (Lonetto *et al.*, 1992), but did not contain the motif (TATAAT) proposed as the consensus -10 vegetative promoter. Box 2 initiates with the sequence ATG; recent-evidence suggests that a 'TG' sequence positioned one base 'upstream' of the σ^{70} -10 hexamer is an essential motif required for transcription and can circumvent the requirement of a -35 element in *E. coli* (Barne *et al.*, 1997).

An over-riding feature of these promoter elements seems to be the extended nature of the -35 and -10 elements. The -35 box is 17 bp in length while and the -10 box extends for 13 bp. If these are the promoters of the neurotoxin operons, the extended lengths recognised may explain the poor H-T-H prediction of their respective transcriptional factors, which may be based on H-T-H motifs recognising shorter DNA sequences (Dodd & Egan, 1990).

The spacing of 12 bp between the promoter elements box 1 and box 2 is uncharacteristic of the σ^{70} family. However, if the sequence 'TTTACA' is used as the core σ^{70} -35 sequence, the distance between this sequence and the -10 box is 15 bp. If the base 3' to the 'TG' motif is used, there is a distance of 19 bp, which exceeds the consensus spacing (17 bp) of a σ^{70} type promoter (Mulligan *et al.*, 1985). As the class of sigma subunit may be largely responsible for -35 to -10 promoter spacing

(Dombroski *et al.*, 1996), the association of an alternative type of sigma factor may suggest that the proposed non-optimal spacing is indeed sigma factor specific.

The spatial association of ORF 3 homologues with the promoter elements box 1 and box 2 in distinct clostridial neurotoxin operons would seem to suggest that their functions are linked. When the H-T-H motif SRQ and possible -35 promoter similarities of the ORF 3 protein, UviA and TxeR are taken together, a convincing hypothesis emerges, relating a group of transcriptional factors to a conserved promoter sequence. However, the -10/box 2 sequences linked to the respective -35/box 1 sequences of the different systems do not appear to correlate (figure 4.8). Examination of the *C. perfringens* pIP404 UV-responsive promoters associated with *uviA* show a striking feature in that the -10 elements associated with promoter sequences TTTACA are unusual (CTTTTAT) and conserved (Garnier & Cole, 1988). Thus, an attractive explanation could be that this class of transcriptional factor recognises specific -35 and -10 elements, thus conferring a high degree of specific control to an operon or regulon. Indeed, the P1, P3 and P5 promoters mapped at the *uviA* locus were not found to be activated *in vitro* by RNA polymerases isolated from *C. perfringens* or *B. subtilis* during vegetative growth (Garnier & Cole, 1988). Furthermore, during an *in vivo* reporter experiment conducted in *C. botulinum* type A strain, ATCC 3502 (Dr I. Henderson, pers comm.), the promoters which were proposed for NTNH-BoNT/A operon and haemagglutinin operon (Henderson *et al.*, 1996) were not active until the stationary phase of growth.

The contribution of additional transcriptional factors such as *Spo0A* are unknown, as are the contribution of the -35 to -10 spacer region which are virtually conserved in the

promoters of the haemagglutinin operon of *C. botulinum* NCTC 2916 and the *C. tetani* neurotoxin.

4.3.4 The NTNH proteins

In *C. botulinum* NCTC 2916, the gene encoding a NTNH protein is found 5' to the respective type A neurotoxin and type B neurotoxin pseudogene. NTNH proteins are found naturally complexed to all known botulinal neurotoxins by non-covalent attractions, resulting in the formation of a 12 S NTNH/neurotoxin complex. *C. botulinum* strains which possess haemagglutinin operons can form larger neurotoxin complexes of 19 S (type A₁) and 16 S (type A₁, B, C, D and G), whilst 12 S complexes are formed by serotypes A₁-D, G and neurotoxic *C. botulinum* strains which do not appear to possess a contiguous haemagglutinin operon (types E, F and A₂).

The molecular dissection of the NTNH genes from various *C. botulinum* serotypes, in combination with a detailed study of the interaction between the components of the neurotoxin complex (Inoue *et al.*, 1996), has improved the understanding of the NTNH gene structure. Alignment of NTNH genes from serotypes A₁, A₂, B, C/D, E and F show that the genes can be sub-divided into two groups, based on those with or without a N-terminal deletion. NTNH genes with an N-terminal deletion belong to serotypes A₂, E and F, and correspond to strains which do not possess contiguous haemagglutinin operons and in which the toxins are only found as a 12 S complex.

Strains without the N-terminal deletion in the NTNH gene, produce either the 12 S or 16 S complex. Inoue has suggested that a protease cleavage site exists in the N-terminus of NTNH proteins which do not have the deletion (A₁, B, C and D), (Inoue *et al.*, 1996). Cleavage at this site destroys a putative haemagglutinin binding site (Fujita *et al.*, 1995)

which blocks further binding of the haemagglutinin components to the 12 S complex, thereby inhibiting the formation of 16 S complex. A 19 S complex is formed by two 16 S complexes, cross-linked by a single haemagglutinin protein (Inoue *et al.*, 1996).

A striking feature of the NTN_H genes from *C. botulinum* NCTC 2916 is the non-uniform distribution of sequence anomalies. Virtually all the differences occur in the 3' 516 bp. This phenomenon seems to occur between other representative NTN_H genes (Kubota *et al.*, 1996; East *et al.*, 1996) and could have arisen by a recombination mechanism. The generally very high relatedness of these genes could facilitate these events.

4.3.5 Recombination

What could be gained through recombination? It must first be noted that the NTN_H gene is not the only site of recombination. The type C and D neurotoxins are preceded by virtually identical NTN_H genes and haemagglutinin operons (Ohyama *et al.*, 1995), which were most likely created from a recombination event involving the entire HA-NTN_H. Hybrid neurotoxin genes also exist which appear to be chimeras of type C and D neurotoxins (Moriishi *et al.*, 1996a and b). Thus, a situation exists where neurotoxin genes, or functional domains of neurotoxin genes, are found in combinations with a variety of nontoxic components. This increased 'repertoire' may be able to increase or alter the environments in which the neurotoxin complex can endure or alter the site of intestinal absorption and so alter pathogenicity. In this light, the NTN_H gene could be a common site at which ancillary operons are matched with different neurotoxins, so called 'module shuffling' (Chateau & Bjorck, 1994).

The mosaic nature of genes and regulons in bacteria is a recurring feature, which is particularly evident in pathogenic organisms, including the vacuolating cytotoxins of *Helicobacter pylori* (Atherton *et al.*, 1995), δ -endotoxins of *Bacillus thuringiensis* spp. (Bravo, 1997) and streptokinase from *Streptococcus pyogenes* (Kapur *et al.*, 1995). This phenomenon probably reflects the importance of horizontal gene transfer and recombination in the evolution of bacterial function and could imply that the neurotoxin complexes of *C. botulinum* are under a degree of selective pressure.

4.3.6 What advantage does a silent neurotoxin gene confer to *C. botulinum* NCTC 2916?

C. botulinum NCTC 2916 had appeared to contain an active type A neurotoxin, a dysfunctional type B neurotoxin and two sets of virtually identical NTN_H and haemagglutinin proteins. There would appear to be little advantage in duplicating these genes. This poses the question, why does this duplication occur? Analysis of the neurotoxin operons from *C. botulinum* 667 (A/B^o) (Hutson *et al.*, 1996) revealed that the silent type B neurotoxin gene was preceded by a NTN_H gene and HA34 gene which were identical to the corresponding locus in *C. botulinum* NCTC 2916. However, the type A neurotoxin gene was accompanied by a NTN_H gene and contiguous operon which resembled the genes found in front of type A₂, E and F strains which did not contain the haemagglutinin operon. Inspection of the NTN_H/A gene from strain 667 reveals two restriction sites (*Xba*I and *Bgl*II) which were predicted to occur 5' to the type A neurotoxin gene in *C. botulinum* NCTC 2916 (Thompson *et al.*, 1990). This would indicate that the clones 5' to the type A neurotoxin were incorrect and were actually derived from the silent type B locus. Furthermore, the PCR cloning strategy used to derive probes for the type A upstream region was designed in an area of the type

A NTN^H gene (clone p933, Henderson *et al.*, 1995) which was essentially identical in both the type A and silent type B loci. The second 5' oligonucleotide primer would have bound to a haemagglutinin derived sequence present only in the silent B neurotoxin locus and not to the p47-like sequence 5' to the type A neurotoxin.

Evidence supporting this hypothesis is presented by East (1996) who demonstrated that PCR oligonucleotide primers which specifically amplified a DNA fragment from NTN^H genes with a N-terminal deletion (typically found in A₂, E or F like NTN^H genes) produced a positive DNA fragment in *C. botulinum* NCTC 2916 and other type A1/B° strains. This mistake would have been circumvented if the restriction profile generated for the type A neurotoxin (Thompson *et al.*, 1990) had not been overlooked.

Elements of this chapter describing the sequence analysis and mRNA transcripts, therefore, apply only to the silent type B neurotoxin locus and would suggest that the regulon is transcriptionally active. In this light, strains containing both type A and silent type B loci would benefit from two sets of nontoxic components, those derived from the type A locus, including NTN^H, and p47 type proteins, similar to the type strain Kyoto F (East *et al.*, 1996), plus the NTN^H and haemagglutinin genes derived from the silent type B locus. This 'choice' may be advantageous to the clostridium in different environments and mechanisms of action (i.e., ingestion of preformed toxin vs. intestinal colonisation) or against different potential victims.

CHAPTER 5

DEVELOPMENT OF A NOVEL SHUTTLE VECTOR CAPABLE OF REPLICATION IN CLOSTRIDIA AND

Escherichia coli.

5.1 Introduction

5.1.1 Why develop another vector system?

The objective was to develop a means of monitoring gene expression at the neurotoxin locus of *C. botulinum* type A using a suitable reporter vector. This genetic system could be achieved by either fusing the reporter gene to the neurotoxin genes of interest, and maintaining the construct on an autonomously replicating plasmid, or by integrating the reporter gene into the gene of interest in the host chromosome. To achieve either strategy, the development of a gene transfer system for *C. botulinum* type A was necessary.

Amongst the clostridia a number of gene transfer systems already exist, principally conjugative plasmids, conjugative transposons, transposons and plasmids. Several features preclude the use of conjugative elements. Firstly the introduction of genes into an organism which may facilitate the inter-species transfer of a virulence factor creates an element of risk and should be avoided. In the same vein, the inclusion of a conjugative origin of transfer i.e, *oriT* from the IncP plasmid RK2 may complement host conjugative systems and facilitate the spread of virulence genes. Furthermore, conjugative elements are generally rather large and possess few unique restriction sites, which makes them less than ideal for molecular manipulations. Transposons are normally maintained within the chromosome and may, therefore, disrupt essential features of the system under study, again making them unsuitable.

Plasmids are relatively small and can be easily engineered to contain unique restriction sites, and thus, they offer a number of practical advantages. Recombinant vectors for use in clostridia have been developed from a number of plasmids of Gram-positive origin including pIM13 from *B. subtilis*, pJU122 from *C. perfringens* and pUB110 from

S. aureus (reviewed by Minton *et al.*, 1993), but on closer inspection, do not necessarily meet the criteria of this project. The majority of these plasmids replicate via the rolling circle mechanism using a single stranded DNA intermediate (Gruss & Ehrlich, 1989). These single stranded intermediates are highly recombinogenic and, therefore, make this type of plasmid intrinsically unstable (Janniere *et al.*, 1990) and would not be expected to efficiently maintain portions of the vector that are not under direct selection. Two vectors which replicate by the unidirectional theta mechanism are derived from the plasmid pAM β 1 (Bruand *et al.*, 1993) and possibly pIP404 (Garnier & Cole, 1988). Plasmids replicating via this method are considered to be more stable. However, when the replicons are isolated from their natural plasmid environments a degree of segregational instability occurs (reviewed by Minton *et al.*, 1993). In summary, no vector currently available completely satisfies the requirements of this project.

5.1.2 Important features in vector design

The ideal *E. coli*-clostridial shuttle vector should comprise a replicon which functions in *E. coli*, a replicon which functions in the clostridial host(s) and markers for selection in both hosts. Group I *C. botulinum* are sensitive to erythromycin (Cato *et al.*, 1986), as are a number of other Gram-positive bacteria. Thus, the *erm* gene from the plasmid pAM β 1 was chosen as a selection marker for clostridia (Brehm *et al.*, 1986). The Tn3-derived β -lactamase was chosen as a selectable marker in *E. coli* (Uhlen & Nordstrom, 1977). The *E. coli* replicon most widely used is that derived from the *E. coli* plasmid ColE1 (Staudenbauer, 1978), because it is stable and can be manipulated to give a range of useful plasmid copy numbers. One such derivative of the ColE1 replicon is found in the pMTL series of vectors (Chambers *et al.*, 1988).

Plasmids which are not shuttle vectors but are designed to only replicate in *E. coli* and serve as a template for homologous recombination in a target host (such as clostridia), are termed “suicide vectors”. Accordingly only an *E. coli* replicon and resistance marker are present on the suicide vector, whilst a second Gram-positive resistance marker is frequently integral to the homologous integration cassette. Irrespective of the use of a vector, plasmid constructs must be genetically manipulated. To achieve this, the necessary restriction endonuclease sites required must be preferably both unique and concentrated in a polylinker which is distinct from functional elements of the plasmid.

In some instances, instability of a vector is thought to be caused by poor recognition of vector replication signals by a heterologous host, for example, the inefficient use of lagging strand conversion signals to convert single stranded plasmid intermediates into double stranded plasmids (Minton *et al.*, 1993). In light of this fact, the ideal choice of vector should be a plasmid isolated from the strain under study. However, cloning and sequence analysis of novel plasmids is demanding on resources and may not yield a suitable replicon.

The solution in this study was to choose the cryptic plasmid pCB102, from another mesophilic clostridium, *Clostridium butyricum* NCIB 7423 (Minton & Morris, 1981, Minton *et al.*, 1993). Although pCB102 has been cloned for some time, very little is currently known about its novel replicon. However, it does function in heterologous clostridial hosts, including *C. beijerinckii*, and has the advantage of being fully sequenced (Dr N. P. Minton, Pers. comm.).

5.2 Analysis of pCB102

5.2.1 Nucleotide sequence analysis

In silico nucleotide sequence analysis of pCB102 reveals the presence of nine potential ORFs (A-I) each potentially encoding proteins of greater than 70 amino acids (figure 5.1), which appear to constitute at least four operons (Minton *et al.*, 1993). Comparisons of the primary amino acid sequence of each predicted encoded protein with the entire SWISSPROT data base (issue 100) revealed significant similarities with other proteins for only the products of ORF A and ORF I (figure 5.2 and 5.3). Polypeptide A, the product of ORF A, exhibits 22.1 % similarity with MobC, a polypeptide of 94 residues encoded by the *E. coli* IncQ incompatibility group plasmid RSF1010. The *mobC* gene forms part of a divergent operon whose function is to provide a mobilisation mechanism for RSF1010 (Scherzinger *et al.*, 1993). MobS, also thought to have a role in the conjugative mobilisation of another plasmid pTF1 from *Thiobacillus ferrooxidans*, forms a third member of this group. Both MobS and MobC have been shown to be essential in the mobilisation process of their respective plasmids (Drolet *et al.*, 1990 and Scherzinger *et al.*, 1992). In RSF1010, MobC has been shown to bind to a larger mobilisation protein MobA, when MobA is found complexed to the substrate *oriT* site. MobC then induces DNA strand melting allowing MobA to effect a single strand cleavage activity, initiating the plasmid mobilisation process (Scherzinger *et al.*, 1992).

Proteins which are similar to the predicted product of ORF I (polypeptide I) are also exclusively of plasmid origin and form an indispensable part of mobilisation machinery. The most closely related of these is ORF1 (for clarity referred to as ORF st-1) from the streptococcal conjugative plasmid pIP501 (Wang & Macrina, 1995) which shares 58.8

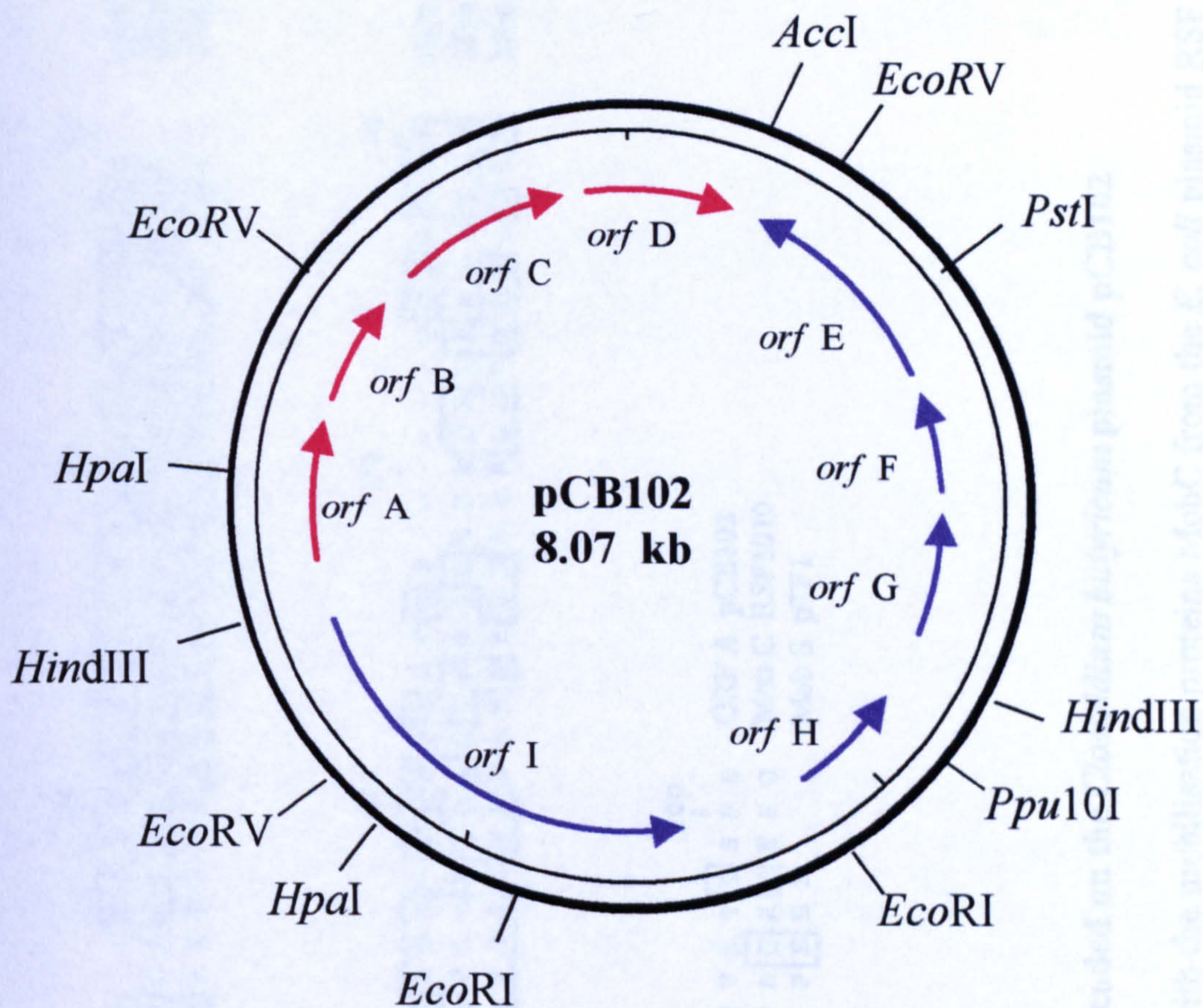


Figure 5.1 Schematic map of pCB102

pCB102 was isolated from *Clostridium butyricum* NCIB 7423 where it was found to be co-resident with a second plasmid pCB101 (Minton & Morris, 1981). (Adapted from Minton *et al.*, 1993).

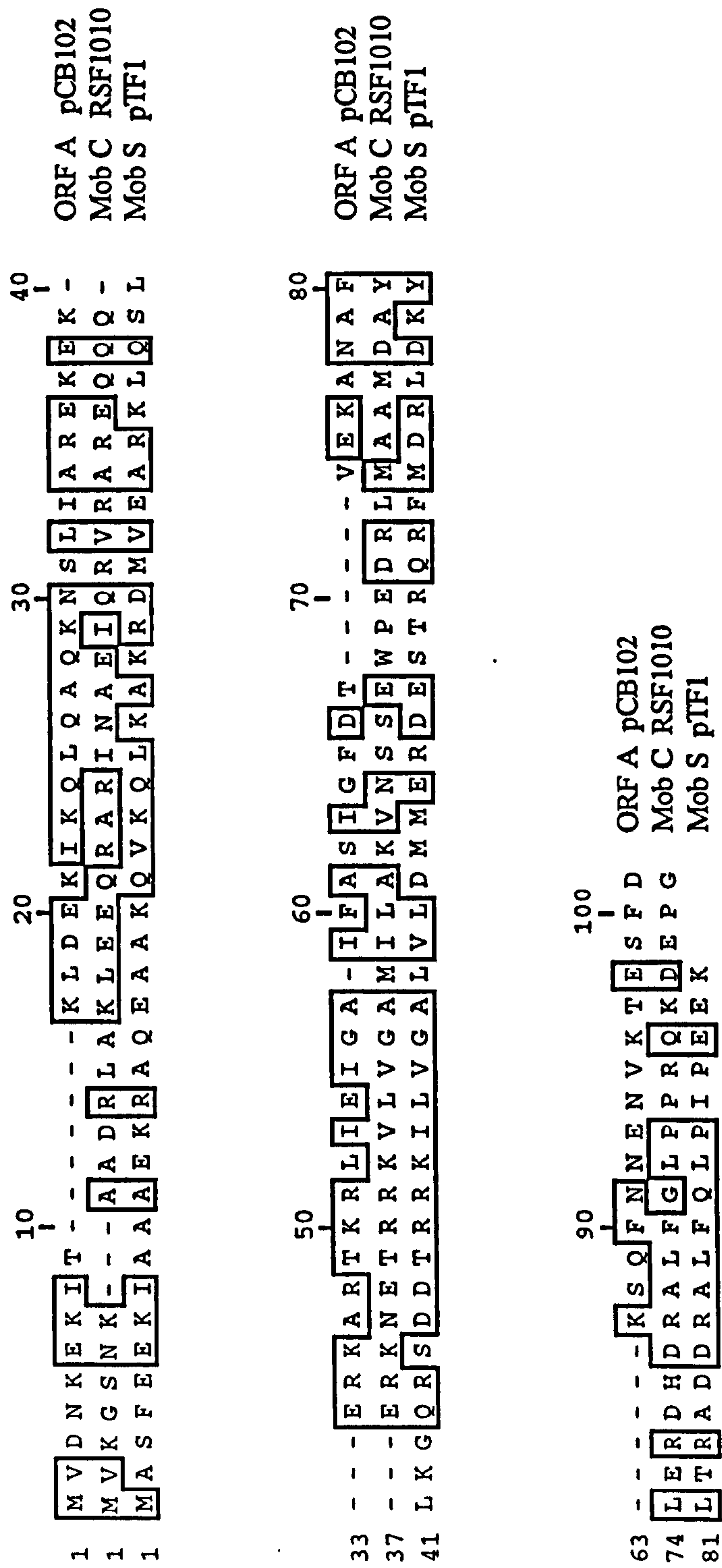


Figure 5.2 Homologues of ORFA encoded on the *Clostridium butyricum* plasmid pCB102

Comparative alignment of ORF A with the mobilisation proteins MobC from the *E. coli* plasmid RSF1010 (Scholz *et al.*, 1989) containing 22.1 % similarity and MobS from a *T. ferrooxidans* plasmid pTF1(Drolet *et al.*, 1990) containing 19.6 % similarity. (Residue groupings are (ST), (KR), (ILV), (EDNQ), (YWF), (A), (G), (M), (H), (C), (P)). The genes of all three proteins are divergently transcribed from a site specific 'nicase' involved in plasmid mobilisation and associated with a conserved *oriT* *nic* site (Drolet *et al.*, 1990).

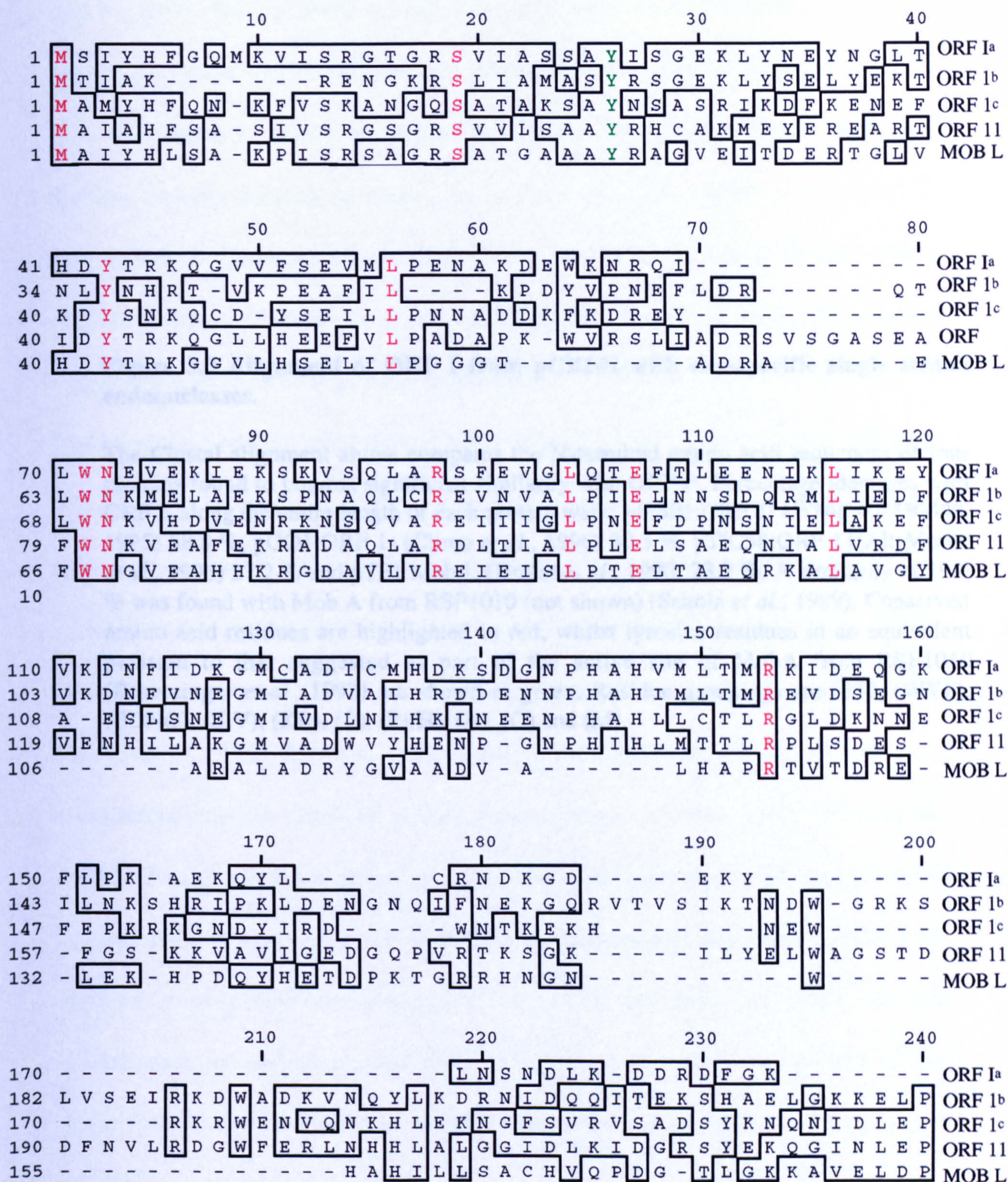


Figure 5.3 Alignment of ORF I from pCB102 with site-specific single strand endonucleases.

Figure 5.3 Alignment of ORF I from pCB102 with site-specific single strand endonucleases.

The Clustal alignment above compares the N-terminal amino acid sequences of four proteins found to possess significant similarity with ORF I^a. Percentage identities with ORF I^a along the entire length of each protein were; pIP501:ORF 1^b (Wang & Macrina, 1995) 58.8 %, pGO1:ORF 1^c (Climo *et al.*, 1996) 30.3 %, pTiC58:ORF 11 (Alt-Morbe *et al.*, 1996) 27.2 % and pTF1:Mob L (Drolet *et al.*, 1990) 23.2 %. A similarity of 19.6 % was found with Mob A from RSF1010 (not shown) (Scholz *et al.*, 1989). Conserved amino acid residues are highlighted in red, whilst tyrosine residues in an equivalent position to that suggested as part of the active site of MobA from RSF1010 (Scherzinger *et al.*, 1993) are shown in green. Residue groupings are (ST), (WYF), (EDNQ), (ILV), (KR), (C), (P)(H), (A), (G) and (M).

% similarity. Polypeptide I and ORF st-1 appear to be members of a much larger group of proteins which include MobA from RSF1010 and whose common function is a site specific DNA cleaving-joining activity associated with plasmid mobilisation. Greatest similarity exists in the N-terminal region, extending for the first 160 amino acids (figure 5.3). This region is thought to contain the site specific cleavage activity, involving a tyrosine residue mapped in MobA to position 24. This residue is conserved in polypeptide I. However, MobA of RSF1010 possess a C-terminal domain with a separate function (Scherzinger *et al.*, 1993), which may reflect the poor sequence conservation between these proteins in this region.

Intriguingly the relative orientation of ORF A/ORF I and their homologs in RSF1010 and pTF1 are conserved in all three examples. Examination of the intergenic nucleotide sequence between ORF A and ORF I revealed the presence of a conserved motif 3' to an inverted repeat, previously identified as an *oriT* in RSF1010 (Scherzinger *et al.*, 1993) and which conforms to the origin of transfer found in the IncQ group of plasmids (Waters *et al.*, 1993). However, polypeptide A homologues are not found in all other mobilisation operons identified in this search (Wang & Macrina, 1995; Scholz *et al.*, 1988). Proteins which are similar to polypeptide I also occur in a 52 kb staphylococcal plasmid pGO1 (Climo *et al.*, 1996), in the conjugative region of plasmid pTiC58 from *Agrobacterium tumefaciens* (Farrand *et al.*, 1996) and with the single stranded endonuclease encoded on plasmid pSC101 (Drolet *et al.*, 1990). All members of this group occur in close proximity to an IncQ type origin of transfer (figure 5.4) (Climo *et al.*, 1996).

pCB102	G	G	G	G	T	A	T	A	A	G	T	G	C	G	C	C	C	T	T	C
RSF1010	A	A	C	C	G	G	T	A	A	A	T	G	C	G	C	C	C	T	C	C
R1162	A	A	C	C	G	G	T	A	A	G	T	G	C	G	C	C	C	T	C	C
pTF1	T	T	A	C	T	C	T	A	A	G	T	G	C	G	C	C	C	T	T	G
pTiC58oriT	C	G	A	G	T	A	T	A	A	T	T	G	C	G	C	C	C	T	T	G
pSC101	A	A	A	G	T	C	T	A	A	G	T	G	C	G	C	C	C	T	G	A
pIP501	T	G	C	G	T	A	T	A	A	G	T	G	C	G	C	C	T	T	A	
pGO1	T	T	C	G	C	A	T	A	A	G	A	G	C	G	C	C	T	T	A	

Figure 5.4 Alignment of the putative origin of transfer from pCB102

Alignment of the sequence located between ORF A and ORF I on pCB102 with the *oriT* regions of other IncQ compatible plasmids (adapted from Climo *et al.*, 1996). Blue letters indicate conserved nucleotides and red letters indicate virtually conserved nucleotides. Vertical arrows indicate mapped cleavage points of the origin of transfer.

5.2.2 Analysis of the pCB102 clone pCB4.

A previous report suggested that a derivative of pCB102, pCB4 (see figure 5.7) is capable of autonomous replication in clostridia (Minton *et al.*, 1993). Inspection of the translated nucleotide sequence of this clone reveals the presence of 2 major ORFs previously assigned H and I. As described in section 5.2.1, the product of ORF I is likely to contain a single stranded endonuclease activity within its N-terminal domain. However the C-terminus bears little similarity to other proteins listed in current data bases, a feature of the other representatives of this group. The C-termini do encode functional domains, for example, a helicase activity is proposed for TraA of pTiC58 (Climo *et al.*, 1996) and a DNA primase activity specific to *oriV* in MobA of RSF1010 and possibly in ORF 1-st of pIP501 (Wang *et al.*, 1995). Thus it is possible that the replication function may (partially) reside in the C-terminus of ORF I. Alternatively the replicative process may be attributed to the product of ORF H and/or hither to unrecognised mechanisms. Therefore, it was necessary to determine the minimal replicon of pCB102 before using it to construct a shuttle vector.

5.3 Determination of the minimal replicon of pCB102

5.3.1 Deletion analysis of pCB4.

The fragment of DNA encoding the essential replication mechanism of pCB102 was established by deletion analysis of pCB4. A series of nested deletions were created in pCB4 (figure 5.7). A deletion from the unique *StuI* and *HpaI* sites was created by digesting pCB4 and re-ligating the blunt-ended DNA strands (figure 5.7), removing 196 N-terminal amino acids from ORF I. A second larger deletion, *StuI* to *LspI*, was made by digesting and modifying the recessed termini created by *LspI* with Klenow fragment,

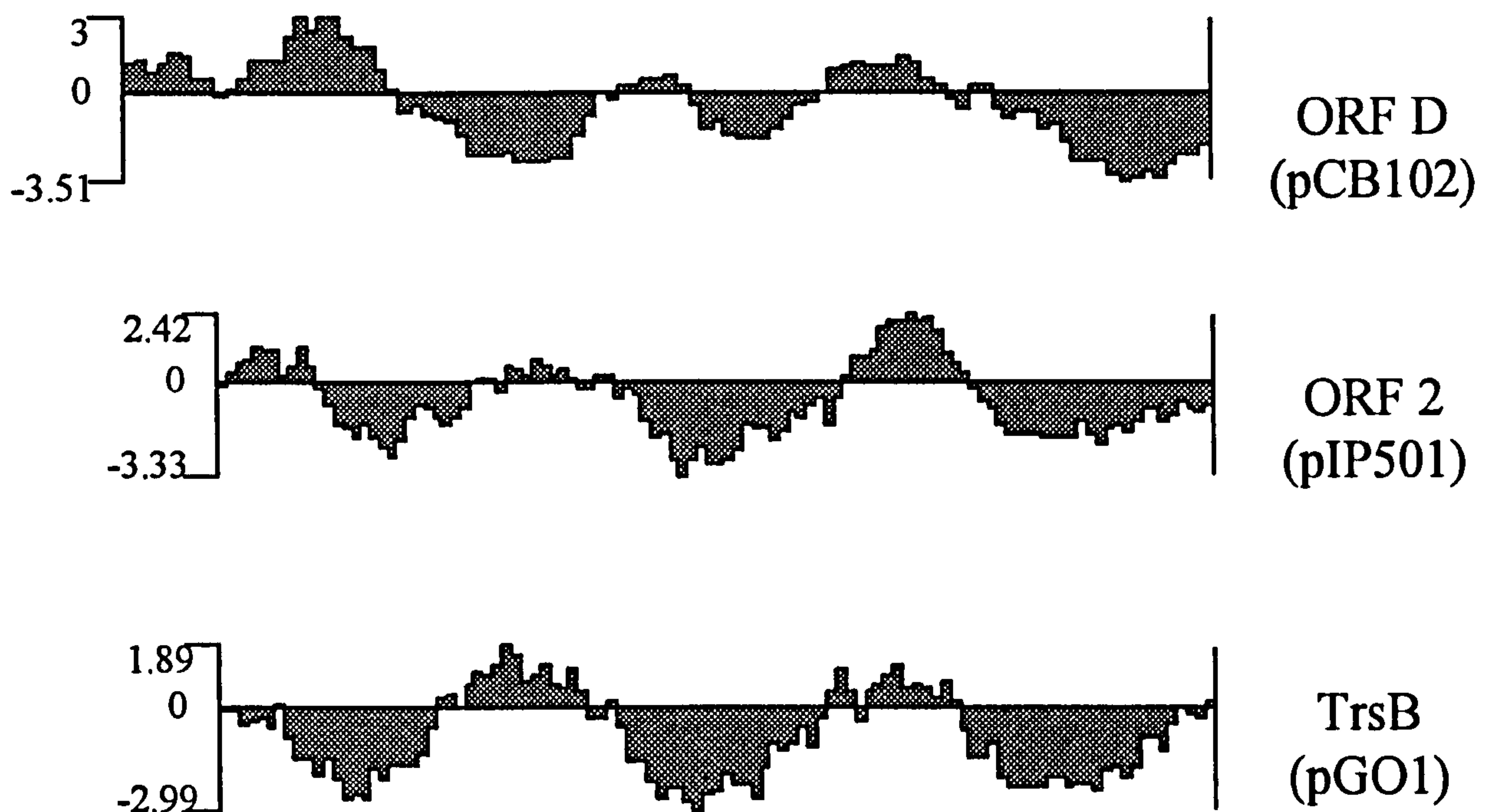


Figure 5.5 Comparative alignment of hydrophilicity profiles for ORF D of pCB102 and putative homologues.

Comparison of the hydropathic profile (Kyte & Doolittle, 1982) for ORF D of pCB102 with ORF 2 from the proposed conjugative operon of pIP501 (Wang & Macrina, 1995) and with TrsB, another member of a Gram-positive conjugative operon (Morton *et al.*, 1993). The pattern of alternating hydrophobic and hydrophilic regions is striking and could suggest that the three proteins may be similar. Hydropathic profiles were predicted using the PROTEAN software package, DNASTAR Inc. with a window of 9 residues. The average hydropathy within this window is represented on a scale where negative values indicate a hydrophobic nature and positive values indicate a hydrophilic nature.



Figure 5.6 Alignment of ORF D (pCB102) with putative homologues

Comparison of the primary amino acid sequence of ORF D from pCB102 with ORF 2 from the broad-host range conjugative plasmid pIP501 (Wang & Macrina, 1995) and TrsB from the conjugative staphylococcal plasmid pGO1 (Morton *et al.*, 1993). Alignment was by the clustal method (Higgins & Sharp, 1986) using the MEGALIGN software package, DNASTAR Inc. with the residue groupings of (ST), (ILV), (KR), (EDNQ), (WYF), (A), (G), (M), (H), (C) and (P). Both ORF 2 and TrsB are found in operons whose protein products are implicated in conjugative mobilisation. In addition all three proteins possess a similar hydrophobicity profile (figure 5.5) suggesting a possible functional and evolutionary relationship.

removing ORF I as well as a further 106 bp of non-coding DNA. The pCB4 derivative plasmids were called pCB4 Δ SH and pCB4 Δ SL respectively (figure 5.7). A third deletion mutant was created by digestion of pCB4 with *EcoRI*, removing ORF I and the native promoter of ORF H. This derivative was named pCB4 Δ RI. Finally, an ORF H frameshift mutant was created by digesting pCB4 with *Ppu10I* (figure 5.7). This restriction site is unusual because when the 3' overhangs created by digestion with *Ppu10I* are filled with Klenow fragment and ligated, two partially overlapping *Ppu10I* sites are recreated. Although *Ppu10I* could not be used to select for frame-shifted sites, the manipulation should have created a new *SphI* site, However no *SphI* sites were detected. Mutant versions were isolated which were refractory to digestion with either enzyme. Subsequent DNA sequencing across the modified site revealed a small 4 bp deletion, not typically associated with the activities of Klenow fragment, but still generating a frame-shift mutation (figure 5.8). This mutant of pCB4 was called pCB4 Δ P.

5.3.2 Functional analysis of pCB4 and its deletion derivatives

The plasmid pCB4 and its deletion derivatives were introduced into *C. beijerinckii* NCIMB 8052 by electrotransformation. Putative transformants were selected on 2xYTG agar supplemented with erythromycin at a concentration of 20 μ g/ml. Transformants were scored after 48 h incubation. Colonies were obtained for pCB4, pCB4 Δ SH, pCB4 Δ SL and pCB4 Δ P; no colonies were obtained with pCB4 Δ RI. It was noted that the number of transformants achieved with pCB4 Δ P was considerably lower (10/ μ g DNA) when compared to that obtained with pCB4, pCB4 Δ SH, pCB4 Δ SL (10³/ μ g DNA each). Plasmid DNA was isolated from clostridial cultures transformed

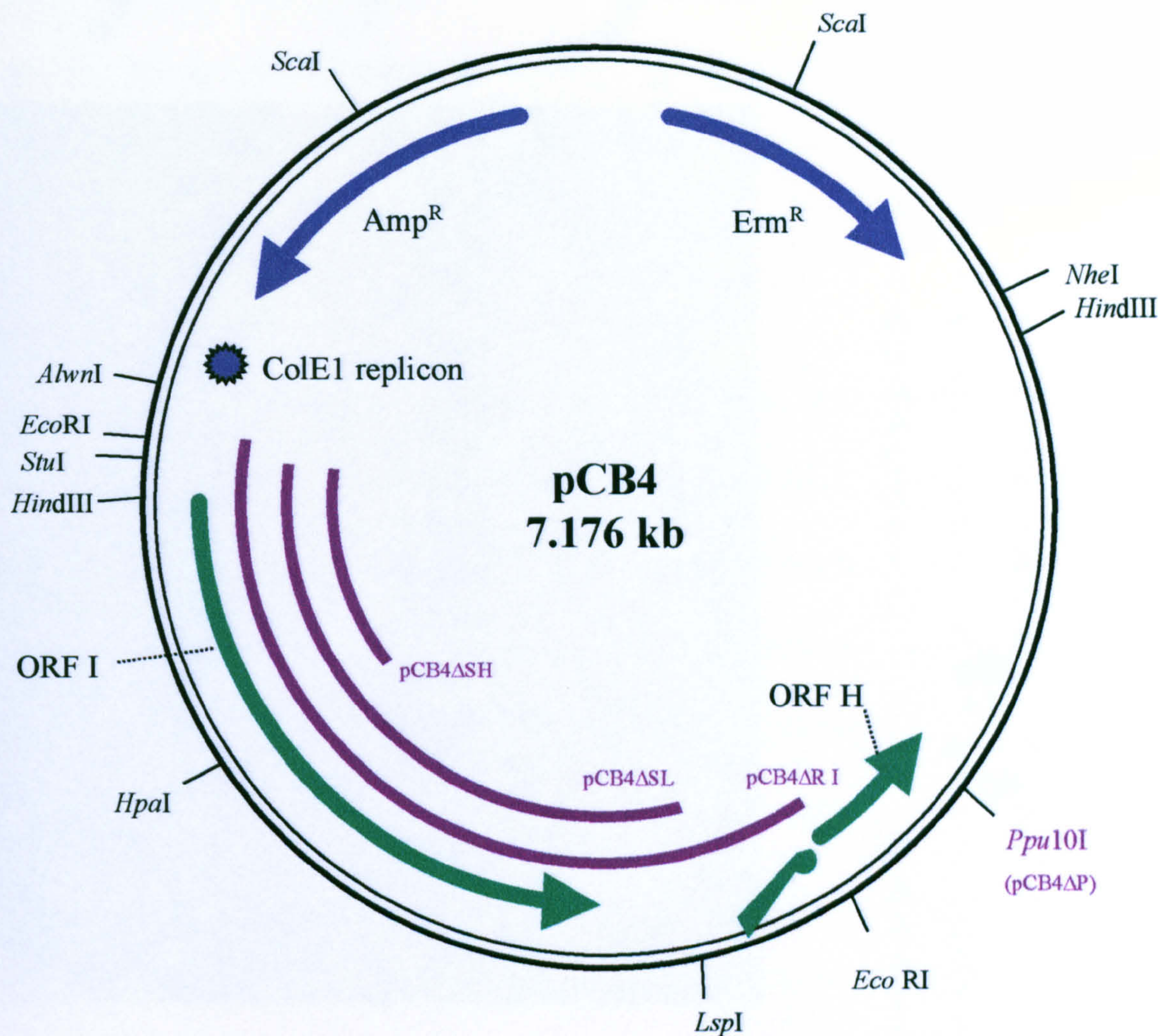


Figure 5.7 Deletion analysis of pCB4

pCB4 contains a 3.6 kb *HindIII* fragment encoding the replicon from a clostridial plasmid pCB102 cloned into a *ColE1* based *E. coli* vector pMTL20E (Minton *et al.*, 1993). Putative ORFs are shown as green arrows. Deletion derivatives (*pCB4ΔSH*, *pCB4ΔSL* and *pCB4ΔRI*) are indicated with purple lines and the frame-shift generated at a *Ppu10I* site (*pCB4ΔP*) is shown in purple.

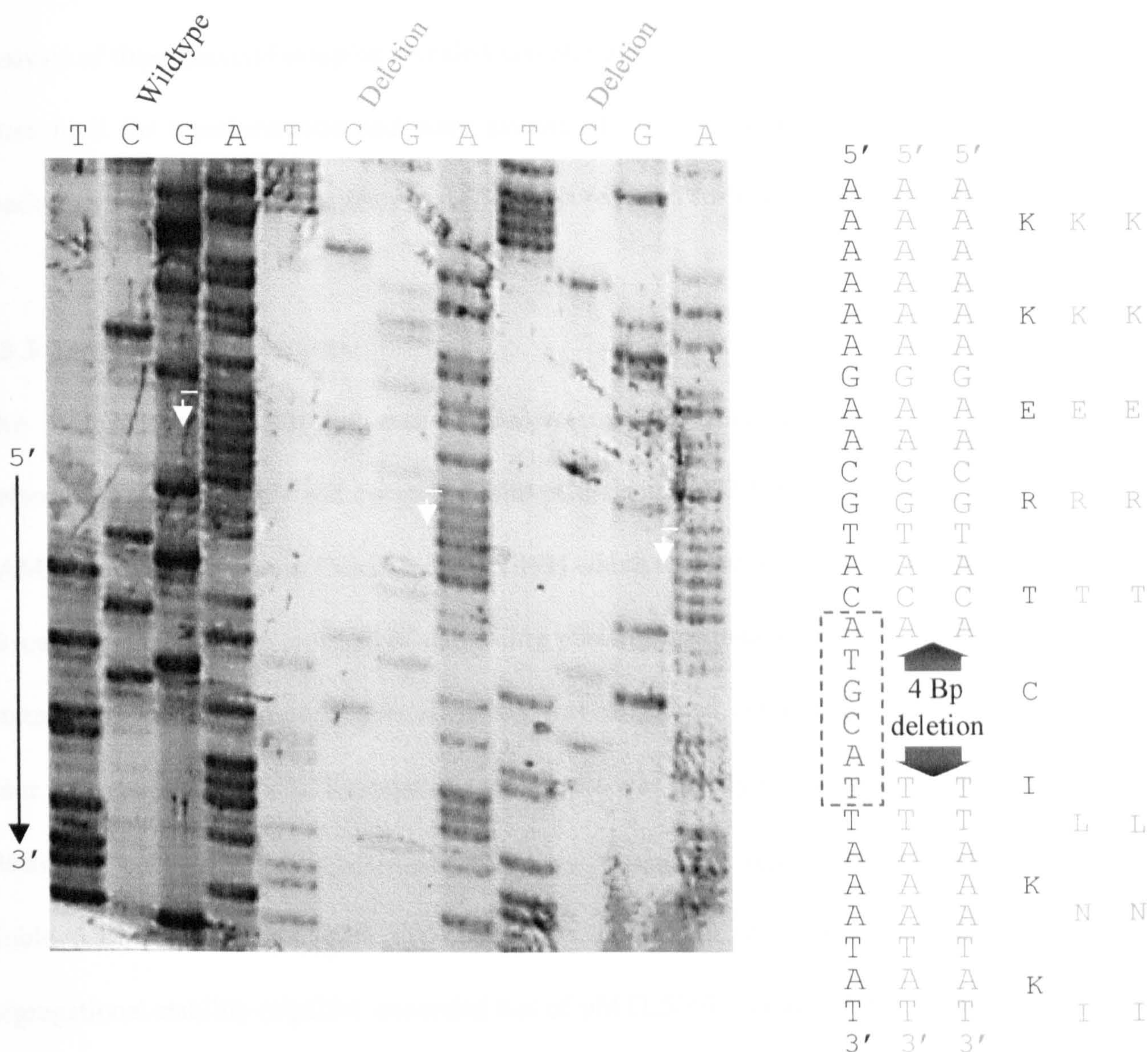


Figure 5.8 Frame-shift mutation of ORF H

The sequence above illustrates a section of ORF H encoded by pCB102 in which a frame-shift mutation was introduced at the *Ppu10* I site (highlighted in red). Wild type sequence is shown in black whereas two separate mutant versions are shown in blue and green. The translated amino acid sequences are shown on the extreme right, corresponding to the central base of the respective codon. A white arrow indicates the position at which the sequence annotation begins.

with pCB4, pCB4 Δ SH, pCB4 Δ SL and pCB4 Δ P. The samples were then introduced into *E. coli* TG1 and a larger quantity of plasmid DNA prepared. Restriction enzyme analysis of these plasmid samples revealed that all constructs appeared to be identical to those used for transformation and were assumed to be structurally stable under the conditions used. This data suggests that ORF H is required for plasmid replication.

5.3.3 Segregational stability.

The segregational stability of each deleted vector construct was estimated in *C. beijerinckii* NCIMB 8052 and compared with pCB4 and pMTL500E, a relatively stable pAM β 1 derivative plasmid (Minton *et al.*, 1993) which replicates via a theta mechanism (Bruand *et al.*, 1993). A method of estimating plasmid maintenance in the multiplying bacterial population (segregational stability) was employed (section 2.2.4). The eight hour time course chosen for inoculating fresh media was strictly adhered to, avoiding the clostridial cells entering a persistent stationary phase. The results are presented in Table 5.1. The data indicates that under the conditions of this experiment, the segregational stability of pCB4 exceeded that of pMTL500E. Indeed, 10 % of the final population had lost pCB4 whilst between 28-74 % had lost pMTL500E. This feature of the isolated pCB102 replicon was extended to the deletion derivatives pCB4 Δ SH and pCB4 Δ SL. In contrast, pCB4 Δ P demonstrated particularly poor segregational stability with 84 % of the population having lost the antibiotic phenotype. It was observed that the degree of segregational stability of a particular plasmid was proportional to the relative electro-transformation frequency among the plasmids tested. The structural integrity of the plasmid constructs was investigated at the end of the stability experiment. Plasmid DNA was purified from replica plated colonies with an

erythromycin resistant phenotype and passaged through *E. coli*, strain TG1. Restriction enzyme analysis of the plasmid DNA revealed no obvious structural alterations in any of the constructs (data not presented).

VECTOR	PCB4	PCB4ΔSH	PCB4ΔSL	PMTL540E	PMTL541E	PMTL500E	PCB4ΔP
Exp.1	92.0	98.0	100.0	92.0	100.0	74.0	-
Exp.2	88.0	98.0	98.0	92.0	100.0	72.0	-
Exp.3	88.0	92.0	96.0	100.0	98.0	34.0	-
Exp.4	84.0	98.0	94.0	100.0	100.0	26.0	16.0
Exp.5	92.0	100.0	98.0	100.0	98.0	26.0	21.0
Exp.6	96.0	100.0	98.0	100.0	100.0	28.0	11.0
Mean	90.0	97.7	97.3	97.3	99.3	43.6	16.0

Table 5.1 Estimation of plasmid segregational stability

Data was obtained by the method described in section 2.2.4. Results are expressed as the percentage of *C. beijerinckii* NCIMB 8052 cells retaining an erythromycin resistant phenotype after 16 h of growth with out antibiotic selection. (-) not done. For plasmids pMTL540E and pMTL541E see section 5.4.

5.4 Construction of an *E. coli*-Clostridial shuttle vector system.

5.4.1 Vector construction

The 1.627 kb (*LspI*-*HindIII*) restriction fragment was isolated from pCB4 and the ends filled in using the Klenow fragment. pMTL21E (see figure 5.11) encoding ampicillin resistance (β -lactamase), an *E. coli* ColE1 replicon and *lacZ* α , was linearised at the unique *NheI* site and the ends filled using Klenow DNA polymerase. The two fragments were ligated to give either pMTL540E or pMTL541E, depending on the orientation of the 1.627 kb of pCB4. The *NheI* sites were recreated at the fusion between the modified *NheI* and *LspI* sites.

5.4.2 Functional analysis of pMTL540E and pMTL541E

In *E. coli*, the plasmid vectors pMTL540E and pMTL541E possessed an ampicillin and erythromycin resistant phenotype, suggesting that both the *bla* and *erm* genes were fully functional. In addition, growth on solid media containing X-Gal and IPTG produced blue colonies, indicating that the *lacZα* component and multiple cloning sites were intact. Electroporation of *C. beijerinckii* NCIMB 8052 with both plasmids resulted in erythromycin resistant colonies. Subsequent passage of plasmid DNA isolated from these colonies through *E. coli* TG1 produced DNA which was indistinguishable from that originally introduced in to *C. beijerinckii* NCIMB 8052. Furthermore, in *C. beijerinckii* NCIMB 8052 the estimation of segregational stability of these constructs in comparison with pCB4 and pMTL500E, demonstrated that pMTL540E and pMTL541E had inherited a degree of segregational stability comparable to pCB4 (table 5.1).

5.5 Construction of a Clostridial *in-vivo* expression system

To develop the range of capabilities for this vector series an expression cassette was incorporated. A simple strategy was employed to replace the *lacZα* component of pMTL540E and pMTL541E with a derivative of the same gene under the transcriptional control of a constitutive Gram-positive promoter. An expression cassette combined with a plasmid encoding the pAMβ1 replicon, termed pMTL500F, had already been constructed (Brehm *et al.*, 1988). This employed the transcriptional machinery of the ferredoxin gene (*fd*) from *Clostridium pasteurianum* (Graves *et al.*, 1985). A *PvuI* restriction fragment encoding the expression cassette was chosen to substitute the original *lacZα* gene of pMTL540E and pMTL541E. Dephosphorylation

was not employed as each fragment involved in the ligation reaction was incapable of autonomous replication. All attempts at this experiment were unsuccessful. A similar approach using the restriction endonuclease *Bgl*I instead of *Pvu*I resulted in the cloning of the correct vectors, designated pMTL540F and pMTL541F (figure 5.12). However, over a number of attempts, only pMTL541F could be introduced successfully into *C. beijerinckii* NCIMB 8052 by electroporation. The presence of a Gram-positive promoter controlling *lacZ* α gene expression was the most significant difference between pMTL540E and pMTL540F. Deletion of the promoter element was achieved by digesting pMTL540F plasmid DNA with *Eco*RV and *Stu*I. This DNA could be used to transform *C. beijerinckii* NCIMB 8052. A number of colonies with an erythromycin resistant phenotype were obtained. The restriction profile of plasmids isolated from these colonies matched the pMTL540F deletion mutant exactly. This experiment suggested that the *Eco*RV-*Stu*I restriction fragment contained a determinant which completely subverted the replicative mechanism encoded by the 1.627 kb pCB102 DNA fragment, in an orientation specific manner.

To test this hypothesis the *fd* transcriptional terminator was inserted between the expression cassette and the replicon. The terminator was amplified by PCR from a genomic clone, pATFdu (Dr J.K. Brehm, Pers. comm). *Xba*I and *Nhe*I restriction endonuclease sites were incorporated into the PCR oligonucleotide primer sequences to facilitate subcloning of the terminator into the unique *Nhe*I site of pMTL540F (figure 5.12). A PCR amplified fragment was initially cloned into a T-tailed vector (section 2.2.12.4) and the nucleotide sequence determined. The terminator fragment was then isolated by digestion with *Xba*I and *Nhe*I restriction endonucleases before ligation into the *Nhe*I site of pMTL540F. Recombinant plasmids were obtained in *E. coli* with a

single copy of the terminator fragment orientated in either direction. The clone with the terminator in the 'natural' direction was designated pMTL540FT, whereas the clone with the alternative arrangement was called pMTL540TR.

Electroporation of *C. beijerinckii* NCIMB 8052 in the presence of either pMTL540T or pMTL540TR resulted in erythromycin resistant colonies. Subsequent analysis of their plasmid content revealed the presence of either vector.

5.6 Discussion

5.6.1 Inspection of the minimal replicon

5.6.1.1 Analysis of ORFs

Deletion analysis of pCB102 revealed that a DNA fragment of 1.627 kb was capable of sustaining replication of plasmid DNA in a clostridial host. This region did not encode ORF I, but was only reported to contain a single ORF, designated H (Minton *et al.*, 1993). When translated, the product of ORF H did not appear to resemble any other protein in the available databases, suggesting that the replicon is novel. The observation that a short deletion near the C-terminus of ORF H (pCB4ΔP) appeared to reduce the plasmids segregational stability seems to implicate this protein in some aspect of replication. A close inspection of the amino acid sequence of ORF H revealed the occurrence of 10 cysteine and 4 histidine residues within a protein of only 98 amino acids. Database analysis of ORF H revealed a low level of similarity with non-structural proteins encoded in rotaviral genomes. Intriguingly, the basis of similarity is with the high content of cysteine and histidine residues (Hua & Patton, 1994). Although no evolutionary relationship is inferred, the function of the rotaviral proteins is thought to be in single stranded nucleic acid binding, associated with viral RNA replication

intermediates. The cysteine rich motifs of rotaviral proteins are thought to form zinc fingers, a motif which also occurs in some transcriptional factors and replication proteins (Coleman, 1992).

An analysis of the DNA sequence between the 5' end of the minimal replicon and ORF H, by translating all six possible reading frames revealed several ORFs greater than 40 amino acid residues, however none appeared to be correlated with convincing initiation codons (ATG, GTG or TTG) or ribosome binding sites (Shine & Dalgarno, 1975). At a distance of 480 bp downstream of the 5' end of the minimal replicon, the DNA sequence was distinguished by the frequent punctuation of all six reading frames with translational termination codons. However, in the first 270 bp directly before ORF H, a short coding sequence of 34 residues is present on the same strand. This feature, named ORF J, begins with a GTG codon and is preceded by a putative ribosome binding site (figure 5.8). Prediction of the secondary structure for this protein revealed a hydrophobic region which could possibly interact with lipid structures such as the cell membrane. In the N-terminal section, a hydrophilic region is attached to the hydrophobic domain with a predicted turn motif (figure 5.13). A second potential ORF (K) exists on the reverse strand overlapping the N-terminus of ORF-H. This 44 amino acid peptide starts with a methionine codon and is predicted to encode three hydrophilic domains interspersed by two hydrophobic domains. Both ORF J and K are reminiscent of small ORFs which are commonly found in the complex control of plasmid replication (Hardy, 1988).

5.6.1.2 Analysis of secondary structure

Computational folding of the entire minimal replication region using the Squiggle program (Zuker, 1989) predicted a number of inverted repeats capable of secondary structure

formation (figure 5.9). A 10 bp inverted repeat was found immediately 5' of the predicted ribosome binding site of ORF H, whilst a second inverted repeat of 8 bp was found 35 bp 5' to the same point. Further upstream, a region of imperfect dyad symmetry was found immediately in front of the putative ORF J. An interesting feature was that the inverted repeat in front of both ORF J and ORF H are nearly identical suggesting that this region could form a much larger secondary structure motif (figure 5.10). This opinion is substantiated by the presence of weaker regions of dyad symmetry in the intervening regions which may form 'lateral appendages' in a larger 'club' structure. It was noted that one short inverted repeat bridged the stop codon of ORF K, a further region of weak dyad symmetry was found 3' to ORF H.

5.6.1.3 Mechanism of replication encoded by pCB102

A previous report has inferred the existence of a single ORF in the replicating region of pCB102 (Minton *et al.*, 1993). Closer inspection has revealed the possibility of a further two significantly shorter ORFs, J and K (Figure 5.9). The existence of short ORFs of 5-11 kDa in close proximity to a replication protein has been documented in plasmids of diverse origins (del Solar *et al.*, 1993; Byeon & Weisblum, 1990; Brantl *et al.*, 1994; Riise & Molin, 1986; Garnier & Cole, 1988). Evidence suggests that several of the proteins have a role in plasmid copy number control. The targets of many of these proteins are nucleic acids, where they recognise inverted repeats. Typically inverted repeats encompass the replication proteins -35 promoter element and therefore exert negative control via transcriptional repression (Brantl *et al.*, 1994). Such copy number control proteins are seldom the only regulatory mechanism. For example, the plasmid replicon of ColE1, which does not encode a 'replication protein,' also possess a small

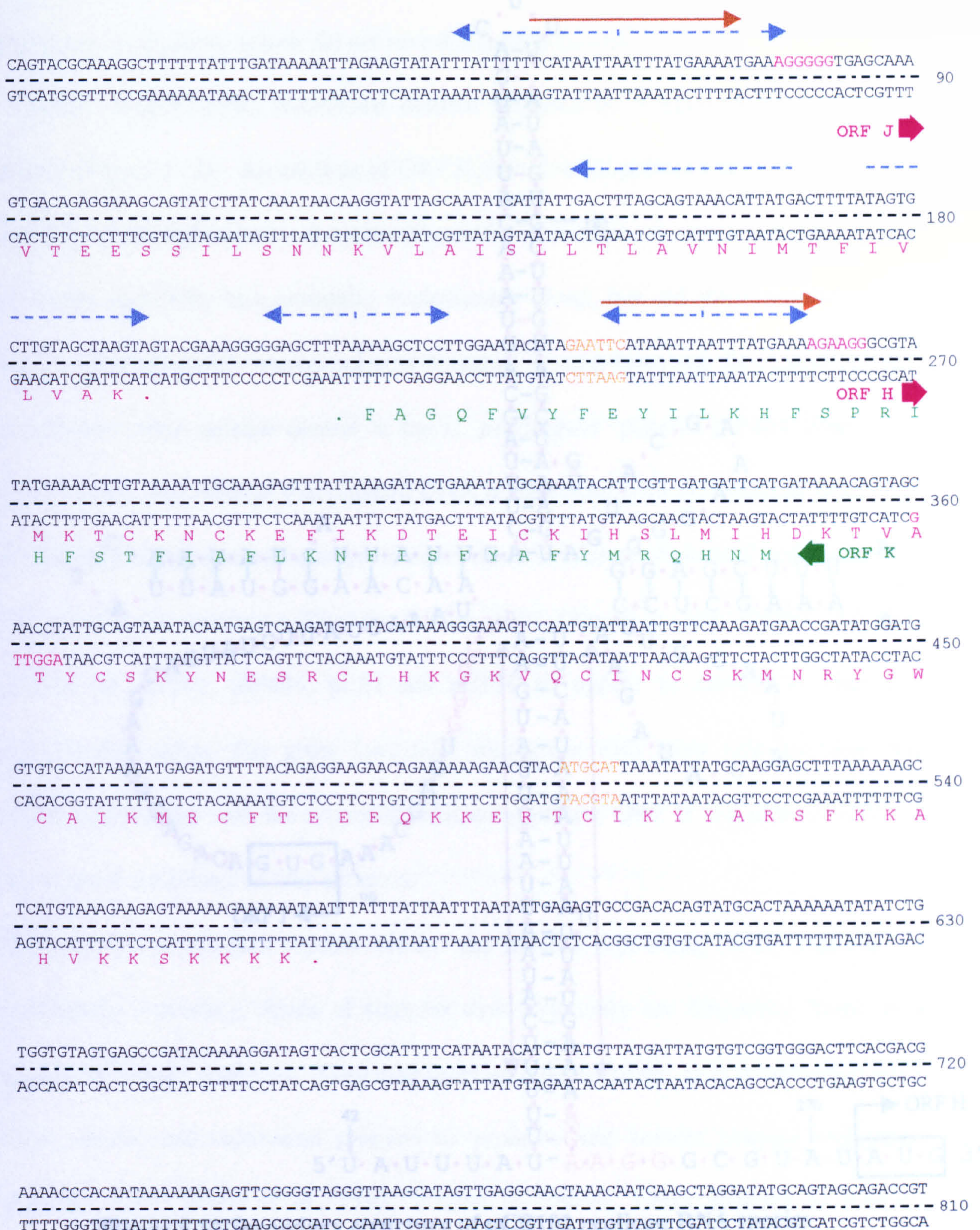
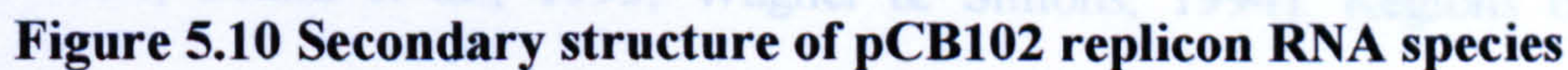


Figure 5.9 Putative regulatory structures within the minimal replicon of pCB102.

Open reading frames are shown below the corresponding DNA sequence. Ribosome binding sites are in magenta. Inverted repeats are indicated with dotted blue arrows, Solid tan arrows denote direct repeats potentially capable of forming a kissing complex. Restriction endonuclease sites, *EcoR* I (GAATTC) and *Ppu*10 I (ATGCAT) are in tan. Note the high occurrence of cysteine and histidine residues in ORF H.



protein of 63 amino acids, 'Rom', which stabilises the interaction of two RNA molecules. A database search did not reveal a close homologue of ORF J, which appears to contain a hydrophilic N-terminal domain followed by a hydrophobic C-terminal domain (Figure 5.13). An analysis of ORF K, revealed the presence of three alternating hydrophilic and hydrophobic domains (Figure 5.13). A database search with ORF K did not reveal similarity to a particular evolutionary group, but did show similarity to a variety of proteins with transmembrane domains. Among these was the protein thought to determine copy number control in the *C. perfringens* plasmid pIP404 (Garnier & Cole, 1988b). The concept that plasmid DNA is associated with the bacterial membrane would suggest the localisation of proteins involved in aspects of plasmid maintenance to be located in the membrane (Firshein & Kim, 1997, Garnier & Cole, 1988b). The Cop proteins of pIP501, pIP404, pLS1 and pE194 all appear to contain at least one hydrophobic region. The gross functional similarities with other Cop proteins and spatial organisation with the proposed replication protein ORF H would make ORFs J and K good candidates as Cop proteins.

The presence of inverted repeats within the minimal replicating DNA fragment is particularly interesting. Areas of complex dyad symmetry are frequently found in a number of diverse replicons. They may function independently, or in combination with, copy number and replication proteins to promote and control plasmid replication (Eguchi, 1991, Brantl *et al.*, 1993, Wagner & Simons, 1994). Regions of complex symmetry may interact at the RNA level. Indeed, pairs of naturally occurring complementary RNA transcripts composing of a sense and antisense strand have been shown to contain sequences capable of forming stem-loop type structures (Wagner & Simons, 1994). Among the potential stem loop structures found in the pCB102

replicon, two appear to occur directly in front of the ribosome binding sites of ORFs H and J (figure 5.9). This relationship of secondary structure and ribosome binding site is encountered in the antisense regulatory mechanism of pIP501 and related plasmids (Brantl, 1993) and the IncFII-like plasmid group (Wagner & Simons, 1994). The nucleotide sequence found in the inverted repeats present in front of the putative RBS of ORF H and ORF J of pCB102 are very similar, sharing 21 bp of homology, interrupted only on one strand by a single base pair insertion. This is reminiscent of structures forming 'kissing complexes' between two complementary stem loop structures. This phenomenon was first characterized in the ColE1 replicon (Tomizawa, 1990) and subsequently in other antisense RNA replicon control mechanisms, e.g., as described by Brantl *et al.*, 1993; Paillart *et al.*, 1996; Eguchi, 1991. An important aspect of the formation of a kissing complex (figure 5.10) is the complementarity between the respective stem structures and the loops. Both requirements are fulfilled by this putative example in pCB102. It is interesting to note that three copies of the stem loop structures could exist, two encoded on a sense strand, as previously mentioned, and one on the antisense strand encoded near the 3' end of the putative ORF K mRNA. At least two further regions of symmetry can be found between the ribosome binding sites of ORF J and ORF H. It is tempting to speculate that they might form part of a larger mRNA secondary structure (figure 5.10) whose formation may also alter the level of translation from ORF H and ORF K ribosome binding sites. The occurrence of an eight base pair inverted repeat overlapping the 3' end of the ORF K coding region could protect this transcript from 3'→5' ribonucleolytic degradation. The half life of a transcript which possibly functions as an antisense RNA would be an important factor in determining copy number control (Brantl & Wagner, 1996).

ORF H would seem to be the candidate replication protein. The evidence in favour is the reduced segregational stability of the *Ppu10I* mutant and that removal of the transcriptional signals of ORF H eliminates the plasmids ability to function. However the matter is complicated by the absence of a designated origin of replication (*oriV*). The replication deficient derivative of pCB4, pCB4ΔR1, which has a deletion of DNA 5' to ORF H, could be due to the loss of a putative *oriV*. Previous work has indicated that a deletion from the *Ppu10I* site to the 3' end of the clostridial DNA results in the total loss of plasmid function. There are two possible explanations, firstly that the *Ppu10I*-*HindIII* fusion (Minton *et al.*, 1993) created a C-terminus which may have been incompatible with the function of ORF H, or that *oriV* was also deleted. *oriV* can occur within the replicating proteins coding region or at either end (Novick, 1989; Garnier & Cole, 1988). Secondary structure predictions for ORF H reveal four alpha helices interrupted by turns. A large proportion of ORF H appears to have the potential to interact with hydrophobic environments, possibly a feature of plasmid maintenance proteins (Firshein & Kim, 1997). The basic nature of the ORF H peptide (pI 9.85) and charge (+17.05) would be consistent with a nucleic acid binding function (Lonetto *et al.*, 1992). Whatever the mode of replication, vectors which are based on the pCB102 replicon would appear to be both segregationally and structurally stable. This would suggest that a single stranded DNA intermediate is unlikely due to the poor stabilities observed in this class of plasmid (Janniere *et al.*, 1990).

The method used to estimate the segregational stability for each plasmid construct was relatively crude. A more accurate experiment should take into account factors such as the relative growth rates of plasmid and non-plasmid containing cells under non-selective conditions or whether the plasmid is conjugative (Boe & Rasmussen, 1996). The data

suggested that very little plasmid loss occurred from generation to generation with vectors based on the pCB102 replicon, whereas a more significant loss occurred using vectors based on the pAM β 1 replicon. However, weight cannot be placed on the degree of segregational stability between vector replicon types, as the relative rate growth rates of plasmid free cells is not known. This fact must also be considered in relation to the pCB4 Δ P stability data. A possible explanation for a plasmid with very low segregational instability could be the consequence of a post segregational killing system. This will not be known until the replicon of pCB102 is analysed in greater detail.

The stability data indicates that pCB4 is marginally less stable than its deletion derivatives pCB4 Δ SH and pCB4 Δ SL. Analysis of the DNA sequence between ORF I and the proposed replicon did not reveal features typical of a Rho-independent transcriptional terminator. Therefore, transcriptional read through from ORF I would seem possible. Transcriptional read through into the replicon could form a natural mechanism of temporary replication attenuation, possibly whilst ORF I and the plasmid are involved in conjugation. This could explain why constitutive expression from the ferredoxin promoter in pMTL540F (figure 5.12) may interfere with plasmid replication.

5.7 Functional deductions for pCB102

5.7.1 Evidence for a Conjugal operon?

pCB102 could be viewed as encoding three separate units, consisting of regulon 1, including ORFs A, B, C, D and I, a second operon encoding ORFs G, F and E, and finally operon 3, encoding ORF H. There is considerable sequence evidence suggesting that regulon 1 may contain features associated with either a conjugal or mobilisable function. Homologues of ORF I and ORF A, in combination with a conserved IncQ

consensus *nic* site, are found in a number of mobilisable plasmids (RSF1010 and pTF1; Drolet *et al.*, 1990) and conjugative plasmids (e.g., Climo *et al.*, 1996). However mobilisable plasmids are distinguished from conjugative plasmids, as they require additional factors donated *in trans* from a truly conjugative plasmid. Intriguingly, a third member of operon 1, ORF D, also appears to belong to a small group of proteins represented in the Gram-positive conjugal systems of pIP501 (Wang & Macrina, 1995) and pGO1 (Morton *et al.*, 1993). A computer aided prediction of the hydrophilicity profile of ORFs B-D suggested that all three proteins exhibited hydrophilic and hydrophobic domains, the latter commonly required in cell membrane complexes. In particular, ORF D displayed an alternating pattern of hydrophilic and lypophilic regions typically encountered in transmembrane proteins (Wang *et al.*, 1995). Similar predictions for ORFs occurring in operons whose proteins are part of a conjugative system demonstrated that two separate Gram-positive plasmids did indeed appear to contain a protein with a similar profile (figure 5.5). All three proteins were close to 11 kDa, had an overall positive charge and a pI of approximately 9. Alignment of their primary amino acid sequence revealed a number of conserved residue groups (figure 5.6), possibly indicating an evolutionary relationship. Spatial linkage of genes with a related function is a common phenomenon, which is frequently observed in other conjugal systems (Wang & Macrina, 1995; Alt-Morbe *et al.*, 1996; Winans, 1996). Inspection of the putative operon 1 structure reveals that ORF B and C form a tightly linked transcriptional unit where the coding regions are separated by a single nucleotide and the putative ribosome binding site of ORF C is located in the coding region of ORF B. At the 3' end of ORF C, a region of dyad symmetry is found. As this occurs in the coding region, it may well confer resistance to 3'→5' ribonucleolytic degradation of the B-C

transcript as this is a feature frequently associated with enhanced gene expression via improved mRNA stability (Belasco & Higgins, 1988). ORFs A, B, D and I are preceded by short intergenic regions which are capable of containing individual promoter sequences, but do not appear to contain regions of dyad symmetry typical of Rho-independent transcriptional terminators. In perspective, the organisation of ORFs A, B, C, D and I may form a divergent operon, and could be designed to produce proteins for the conjugal mechanism at specific ratios. In a comparison with characterised conjugal operons (Wang & Macrina, 1995, Winans *et al.*, 1996, Morton *et al.*, 1993) regulon 1 of pCB102 would contain the least number of genes. This may reflect a narrower host range, or a different membrane complex through which the DNA is transferred.

It is interesting to note that among the large number of single-stranded DNA plasmids characterised, a conjugal system does not appear to have been found. Single-stranded DNA plasmids are thought to mobilise by forming co-integrates with other plasmids using either the RS_A-Pre or RS_B systems (Novick, 1989). Several examples of single-stranded DNA plasmids with a mobilisation system have been investigated (Thomas *et al.*, 1992, Projan *et al.*, 1988). The systems have two *trans*-acting mobilisation proteins and an *oriT*, Transfer must be mediated by the conjugal plasmid pGO1. However no similarities with the *oriT* (IncQ type) specific single-stranded endonucleases were found. A degree of similarity was found with a single-stranded plasmid encoded ORF and an ORF in one of the pGO1 transfer regions (Morton *et al.*, 1993). It would be tempting to speculate that IncQ type *oriT* sites, in combination with their complementary mobilisation proteins and conjugal proteins which may appear to occur in pCB102, were only found in non-single stranded replicating plasmids.

The precise deliniation of this conjugal transfer system is important in pCB102 so the construction of a non-conjugal recombinant vector can be rationally achieved. In addition it seems that plasmids with a full conjugal system are a likely source of a non-single stranded /rolling circle replicating replicon (Bruand *et al.*, 1993). An important factor in vector design.

In summary, the minimal replicon of pCB102 has been mapped to a 1.627 kb DNA fragment which has formed the basis of novel shuttle vectors for members of the clostridia. These plasmids appear to be stably maintained in a heterologous host *C. beijerinckii*, advocating their assessment in other clostridia including *C. botulinum*.

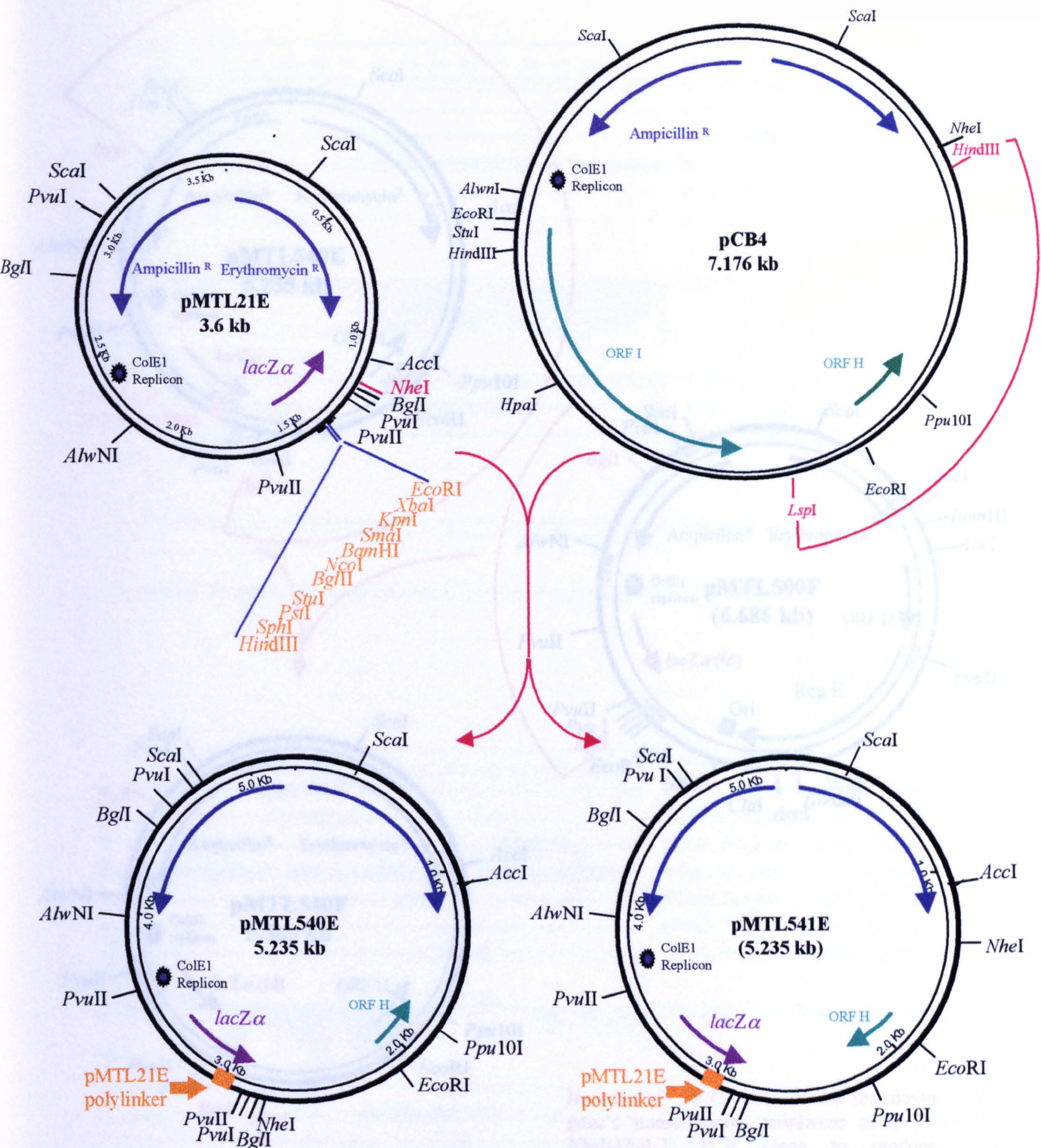


Figure 5.11 Construction of *E. coli*-clostridial shuttle vectors

A versatile pair of vectors encoding both *E. coli* and clostridial replicons were constructed. A blunt ended *LspI*-*HindIII* restriction fragment originally derived from pCB102 (Minton & Morris, 1981) was inserted in both orientations into the modified *NheI* restriction site of pMTL21E (Minton *et al.*, 1993). Vectors pMTL540E and pMTL541E both possess blue-white cloning selection in *E. coli* which is complemented with a versatile polylinker region (shown in tan), plus β -lactamase and erythromycin resistance in Gram-negative and positive organisms respectively.

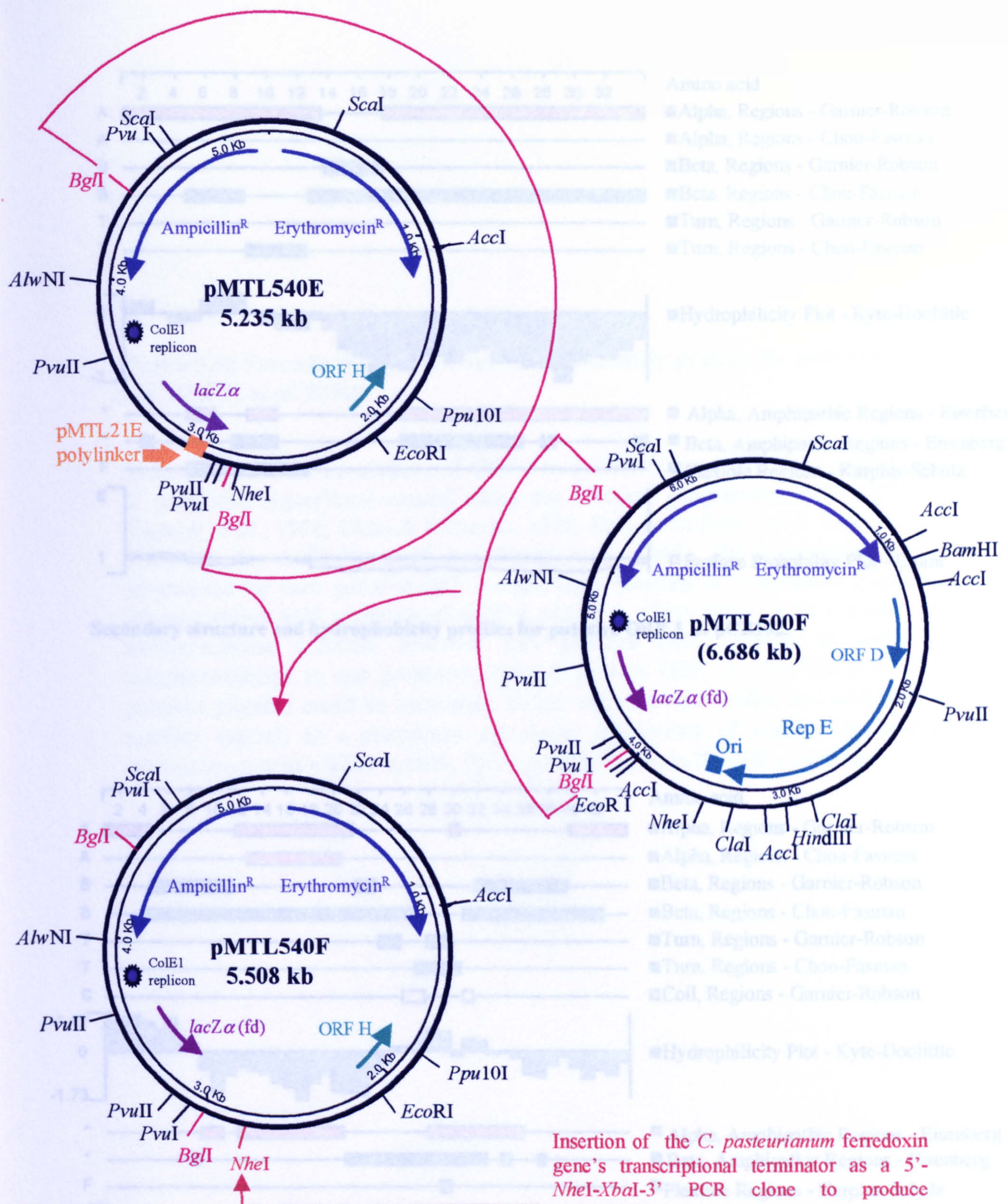
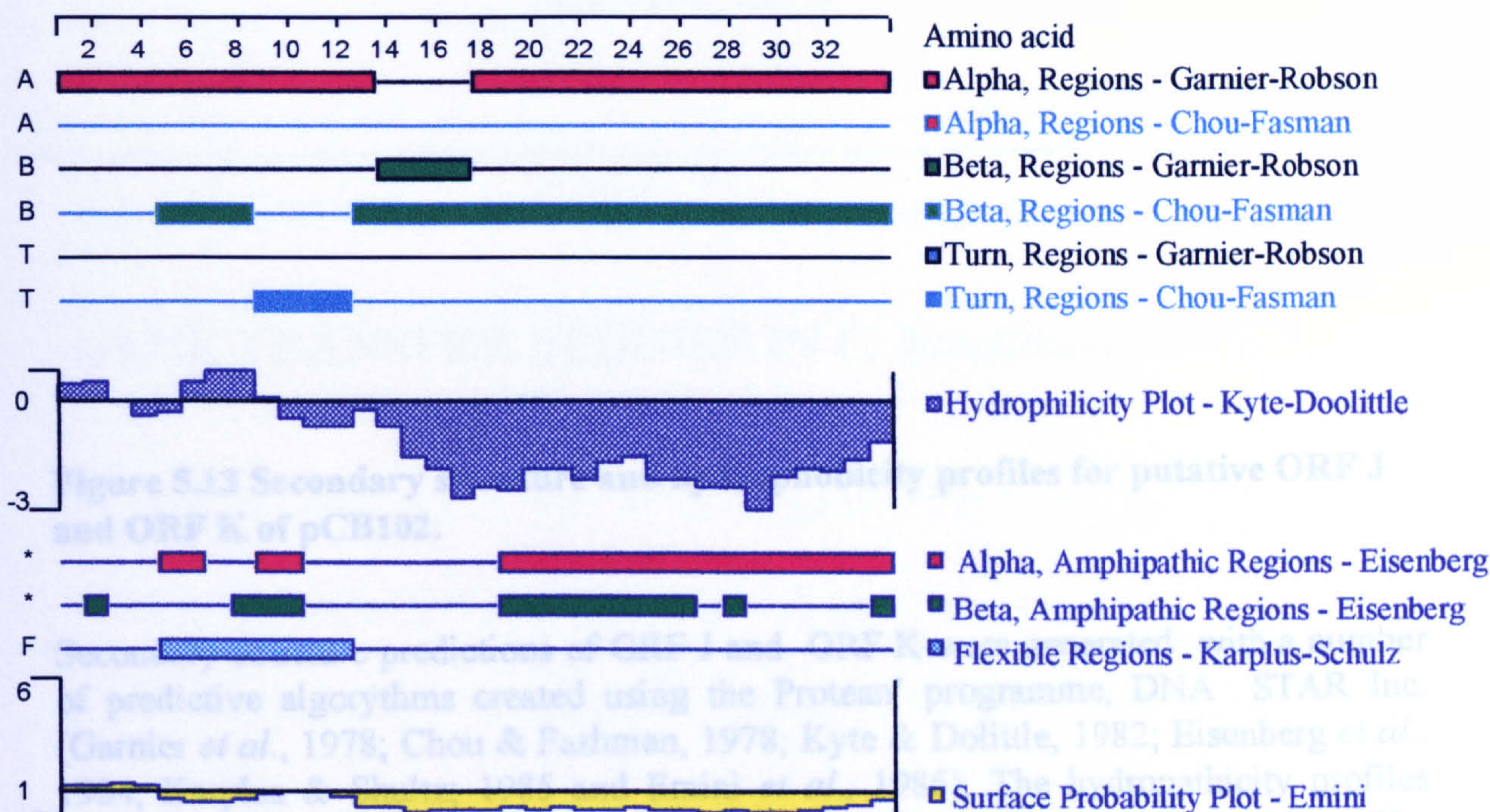
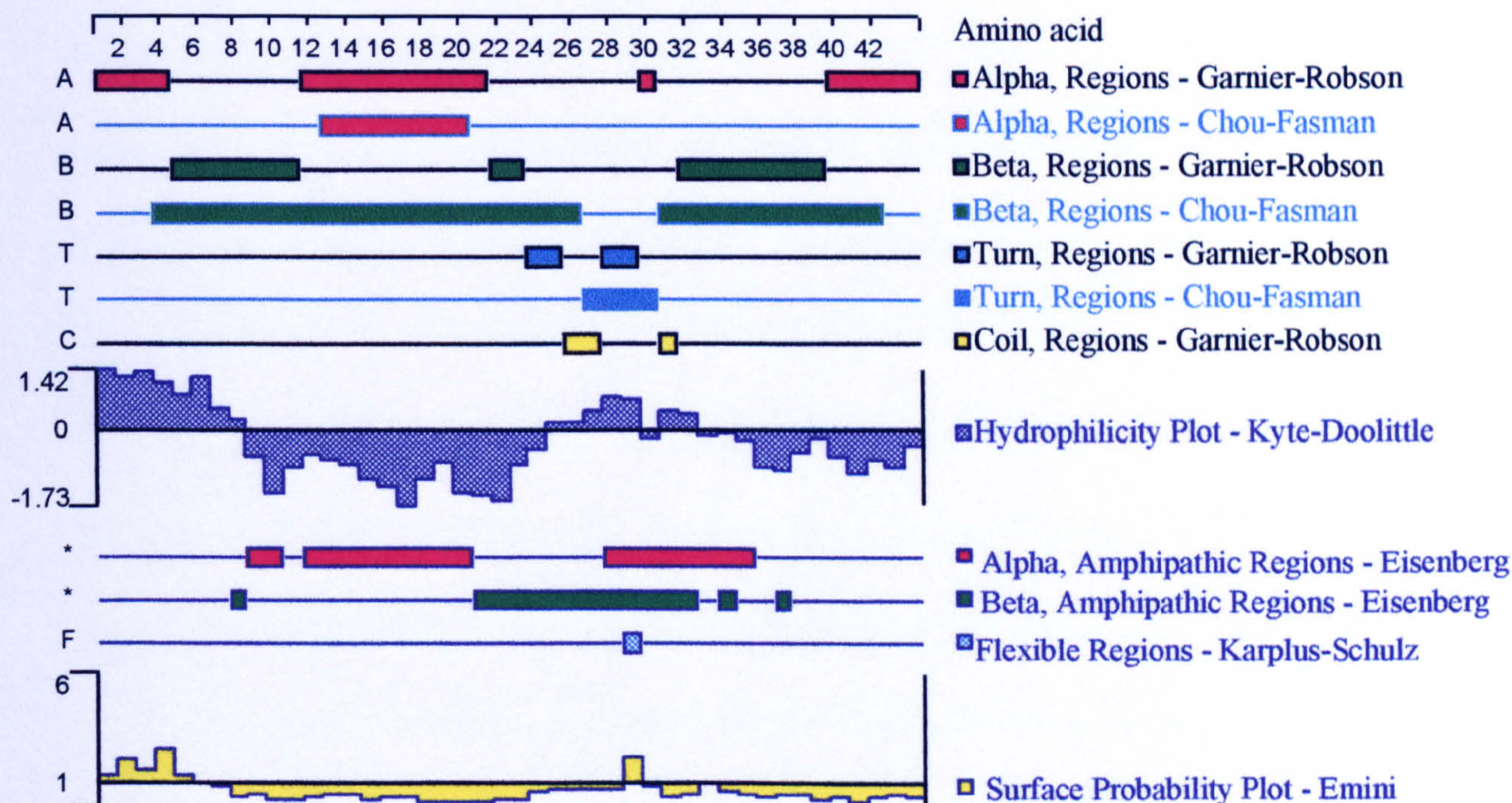


Figure 5.12 Construction of an expression vector system

A Gram-positive expression cassette based on the *C. pasteurianum* transcriptional machinery (Minton *et al.*, 1993) was inserted into pMTL540E by substituting a BglII fragment from pMTL500F. An identical strategy was used to create an expression version of pMTL541E, called pMTL541F. The insertion of a transcriptional terminator into the NheI site of pMTL540F restored the replicative function of the derivative plasmid (pMTL540T) in the clostridial host. Both PMTL540T and pMTL541F contain a convenient array of restriction endonuclease sites in which gene fusions can be made.



Secondary structure and hydrophobicity profiles for putative ORF J of pCB102.



Secondary structure and hydrophobicity profiles for putative ORF K of pCB102.

Figure 5.13 Secondary structure and hydrophobicity profiles for putative ORF K and ORF J of pCB102.

Figure 5.13 Secondary structure and hydrophobicity profiles for putative ORF J and ORF K of pCB102.

Secondary structure predictions of ORF J and ORF K were generated with a number of predictive algorithms created using the Protean' programme, DNA STAR Inc. (Garnier *et al.*, 1978; Chou & Fasman, 1978; Kyte & Doolittle, 1982; Eisenberg *et al.*, 1984; Karplus & Shultz, 1985 and Emini *et al.*, 1985). The hydropathicity profiles determined for each putative ORF contain large sections of hydrophilic residues. The primary amino acid sequence of ORF K shares a short motif with a wide range of transmembrane proteins, however this putative ORF appears to reside as a countertranscript to the proposed replicon protein ORF H. The function of these putative proteins could be intimately linked with plasmid replication such as in copy number control, as a membrane associated mechanism of equipartition in plasmid maintenance or as a killer system, forming lethal pores in the cell membrane.

CHAPTER 6

GENE TRANSFER STUDIES IN *C. botulinum* TYPE A

6.1 Introduction

6.1.1 The objective of gene transfer

The ability to manipulate microorganisms *in vivo* adds another dimension through which they can be studied. A number of powerful techniques including 'gene knock out' mutagenesis, insertional mutagenesis, *in vivo* expression and promoter-reporter studies rely on the ability to introduce DNA into the cell. However, the strategy may not only be dependent on gene transfer, but on the efficiency with which the cells can be transformed. For instance, to generate gene knock outs by homologous recombination, this relatively inefficient process must be counteracted with the efficient delivery of the recombinant gene knock out cassette. Therefore, the organism must ideally be transformable at high frequency, with a range of 'genetic tools'.

In contrast to the genetic paradigms, such as *E. coli* and *B. subtilis*, members of the genus *Clostridium* are in general, poorly amenable to genetic manipulation. Amongst the clostridia, the most significant advances have been made with *C. perfringens*, a medically important pathogen of both humans and animals, and with various strains of solvent producing saccharolytic clostridia such as *C. acetobutylicum* and *C. beijerinckii* (Minton *et al.*, 1993).

A degree of progress has been achieved with the introduction of plasmid DNA (Zhou & Johnson, 1993) and conjugative transposons (Lin & Johnson, 1991) into a few proteolytic strains of *C. botulinum*. However, the plasmid employed (Zhou & Johnson, 1993) propagates via the production of single stranded DNA intermediates and uses a rolling-circle mode of replication (Leenhouts *et al.*, 1991). Such plasmids are intrinsically unstable (Janniere *et al.*, 1990; Kiewiet *et al.*, 1993), and thus, demonstration of plasmid replication in *C. botulinum* with a vector which was considered to be

relatively stable would be a significant advantage, particularly for applications such as promoter-reporter gene fusion technology for regulatory studies.

The aims of the studies detailed in this chapter were to initially demonstrate plasmid-based gene transfer into a group I, type A *C. botulinum* strain, to demonstrate gene transfer with a derivative of pCB102, which does appear to be stable (section 5.4) and to attempt to optimise the frequency with which these strains were transformed.

6.2 Electrotransformation of *C. botulinum* NCTC 2916

6.2.1 Electrotransformation

Initial attempts to introduce a plasmid into a type A *C. botulinum* strain were based on the report of Zhou & Johnson, (1993) who used a 'Hall A' strain. As there are several type A strains under this classification, the "Hall strain" NCTC 2916 was the initial choice as it was also the focus for other genetic investigations in this study.

The conditions of electroporation determined as optimal by Zhou & Johnson (1993) were also used. Cells (0.8 ml of cell suspension was used per cuvette, equivalent to 4×10^8 CFU/ml) were harvested from a TPGY broth culture (OD_{660} 0.8), washed and resuspended in chilled 10 % (w/v) PEG before electroporation in 0.4 cm gap cuvettes at 25 μ F and 2.5 kV. Freshly transformed cells were recovered and plated according to the same protocol. The plasmid used by Zhou & Johnson, pGK12 (see table 2.1), was prepared by the CsCl density gradient method (section 2.2.2.3) and used at 1 μ g/0.8 ml of cell resuspension. The conditions described by Zhou & Johnson, (1993) were successful, but the total number of transformants was low. Typically, between 0 and 10 erythromycin resistant transformants were obtained per 0.8 ml of cell resuspension. Frequently no transformants were detected.

6.2.2 Validation of putative transformants,

Transformants were selected on TPGY agar plates (Appendix II) supplemented with 5 µg/ml of erythromycin at 35°C in an anaerobic environment. After 48 h incubation colonies were observed which continued to 'swarm' laterally, engulfing the entire agar plate. These colonies represented true transformants in comparison to colonies which took a minimum of a week to become visible, were considerably smaller and had a less aggressive swarming phenotype.. Although both types of erythromycin resistant colonies were able to grow on TPGY-agar plates supplemented with erythromycin (>50 µg/ml), growth of the smaller colonies was significantly slower in comparison to the larger colonies and did not result in 'plate engulfment'. Only large colonies could grow in TPGY broth supplemented with erythromycin (>20 µg/ml).

The cell morphology and Gram status of putative transformants were determined by Gram staining. Transformants were also shown to harbour the type A neurotoxin gene by PCR, using oligonucleotides designed to amplify a DNA fragment between nucleotides 180-783 of the type A neurotoxin's gene coding sequence (Dr S. M. Whelan, Ph.D thesis, p181).

6.2.3 Isolation of plasmid DNA from putative transformant colonies

To verify the authenticity of *C. botulinum* transformants, plasmid DNA was isolated from TPGY broth cultures supplemented with erythromycin. Initial attempts at plasmid isolation using a published method (Strom *et al.*, 1984) were unsatisfactory due to the presence of partially degraded chromosomal DNA, and the material obtained could not be used to successfully reintroduce plasmid pGK12 into *E. coli* strains. An alternative method, based on the alkaline lysis procedure (Ish-Horowicz & Burke, 1981;

section 2.2.2.7), did enable small quantities of plasmid to be isolated, which were reintroduced into *E. coli* and subsequently isolated and inspected by restriction endonuclease mapping. These plasmids were shown to be identical to pGK12 control digests (data not presented).

6.2.4 Large scale plasmid isolation from putative transformant colonies

To isolate larger quantities of plasmid DNA, the procedure of Ish-Horowicz & Burke (1981) was adapted to process the larger cell mass. All solutions were increased in proportion to the culture volume. In addition the sample was subjected to a further isopycnic centrifugation step (section 2.2.2.4) to remove any contaminating traces of chromosomal DNA, RNA and protein.

Inspection of the plasmid sample separated in a 0.8 % (w/v) agarose gel indicated that at least 5 species of plasmid existed. This type of distribution is not unusual for a single plasmid, due to the presence of different molecular forms, such as multimeric, relaxed and supercoiled plasmid conformations. Restriction endonuclease analysis of the plasmid sample with enzymes cutting in pGK12, including *Hpa*II and *Sac*II, demonstrated that a specific population of the plasmid DNA did conform to the pattern expected for pGK12. In addition, a quantitatively larger amount of digested DNA fragments migrated in an agarose gel at alternative molecular weights. Such an observation could either have been due to the presence of derivatives of pGK12, or from a novel plasmid(s) indigenous to this strain of *C. botulinum*. As there were restriction endonuclease sites present in the unknown DNA which were not present in pGK12 (including *Eco*RI, *Hind*III and *Bgl* II), and the alternative DNA could be resolved to a single 6.0 kb fragment (figure 6.1), it was concluded that the DNA

complement of this transformed strain at this point consisted of a small indigenous plasmid (≈6.0 kb) and the recombinant plasmid vector pGK12 (4.4 kb).

6.2.5 Optimisation of gene transfer in *C. botulinum* NCTC 2916

6.2.5.1 Oxygen contamination

Before an investigation of 'biological' parameters was conducted, the electroporation method was assessed. As *C. botulinum* is an obligate anaerobe, contamination of cells by any solution containing oxygen was undesirable. To assess the steps of the electroporation protocol which were conducted outside of the anaerobic environment, e.g., centrifugation and electroporation, anaerobic indicator solutions (Don Whitely Scientific Ltd) were aliquoted into both the 250 ml Sorval Dry Spin™ centrifuge tubes and electroporation cuvettes. The centrifuge tubes appeared to eliminate aerobic contamination during a simulation of their use, however, obvious contamination of the electroporation cuvettes was detected even with the addition of sealing caps. In consequence, the entire electroporation apparatus was placed within the anaerobic cabinet for further use.

Oxygen contamination of media and the electroporation buffer was evaluated by the addition of Resazurin, a redox indicator. All substances tested appeared free of oxygen after 48 h within the anaerobic cabinet. Thus, in the following experiments a period of 48 h was the duration used to equilibrate buffers and media prior to electroporation.

6.2.5.2 Estimation of cell viability

The report of Zhou & Johnson (1993) demonstrated that 4×10^8 CFU/ml of *C. botulinum* cells were required in the electroporation to obtain 10^3 transformants /μg of DNA. No

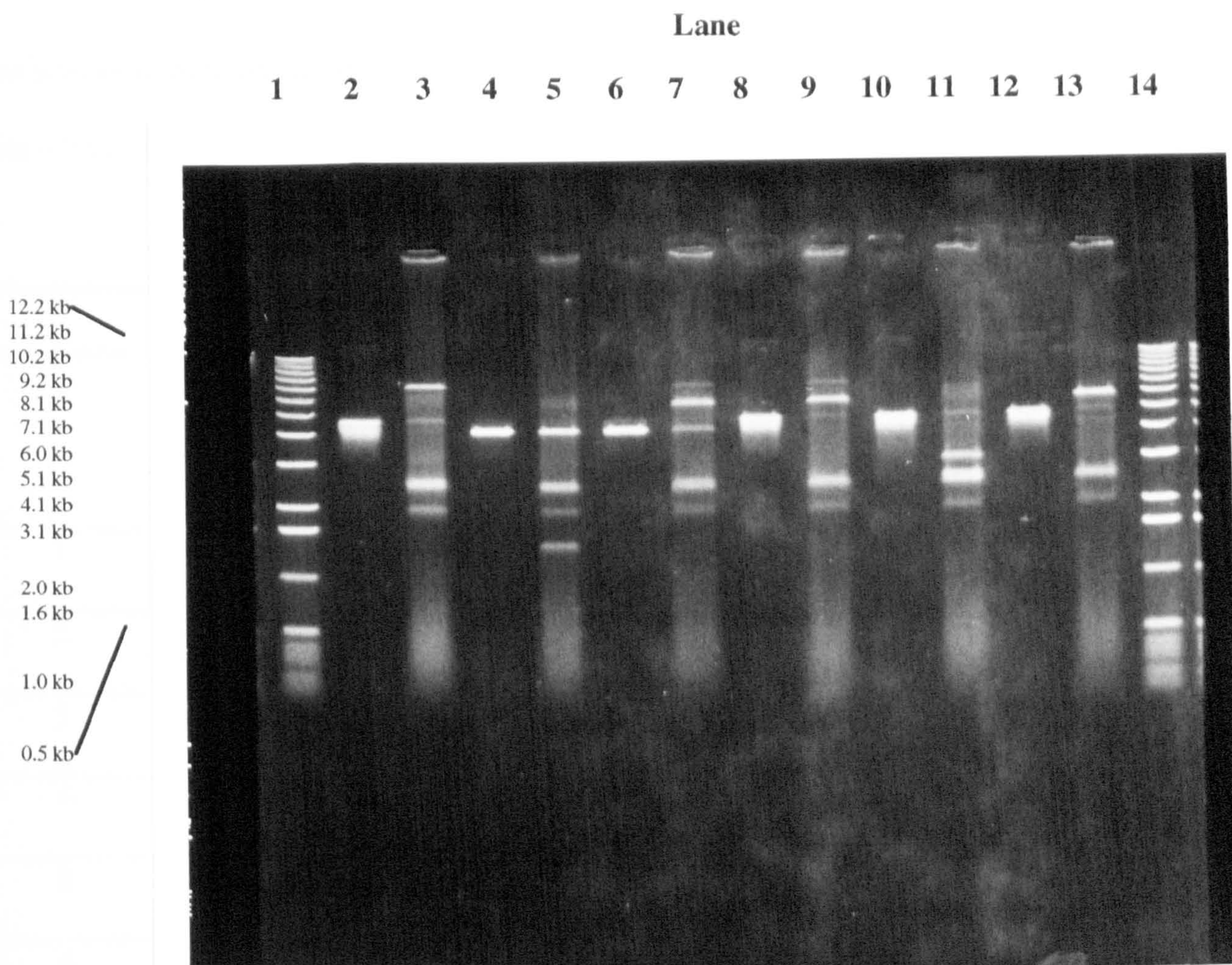


Figure 6.1 Restriction endonuclease analysis plasmid DNA isolated from *C. botulinum* NCTC 2916

Plasmid DNA was isolated from an erythromycin resistant clone of *C. botulinum* NCTC 2916 which was transformed with the recombinant vector pGK12 (Kok *et al.*, 1984). Restriction endonuclease analysis of pGK12 plasmid DNA isolated from *E. coli* (pGK12^{EC}) or *C. botulinum* NCTC 2916 (pGK12^{CB}) demonstrated that in addition to pGK12, a second plasmid of approximately 6.0 kb is present in the clostridial extract.

- | | |
|---|--|
| (1) 1 kb ladder | (8) pGK12 ^{EC} / <i>Eco</i> R I |
| (2) pGK12 ^{EC} /uncut | (9) pGK12 ^{CB} / <i>Eco</i> R I |
| (3) pGK12 ^{CB} /uncut | (10) pGK12 ^{EC} / <i>Hind</i> III |
| (4) pGK12 ^{EC} / <i>Hpa</i> II | (11) pGK12 ^{CB} / <i>Hind</i> III |
| (5) pGK12 ^{CB} / <i>Hpa</i> II | (12) pGK12 ^{EC} / <i>Bgl</i> II |
| (6) pGK12 ^{EC} / <i>Sac</i> II | (13) pGK12 ^{CB} / <i>Bgl</i> II |
| (7) pGK12 ^{CB} / <i>Sac</i> II | (14) 1 kb ladder |

details of culture volumes were given. To correlate the transformability of *C. botulinum* NCTC 2916 with the number of viable cells, electroporation was repeated as described in section 6.2.1, except that cell samples were diluted in electroporation buffer. CFU's were estimated from the original cell resuspension prior to electroporation by generating serial dilutions which were plated onto non-selective media.

Dilution number	Dilution cell:buffer	Estimated CFU/ml experiment 1	Number of transformants (Em ^R) experiment 2	Estimated CFU/ml experiment 1	Number of transformants (Em ^R) experiment 2
1	-	6.5x10 ⁷	-	1.19x10 ⁸	-
2	-	6.5x10 ⁷	12	1.19x10 ⁸	24
3	1:1	3.3x10 ⁷	24	6.0x10 ⁷	24
4	1:3	1.6x10 ⁷	12	3.0x10 ⁷	18
5	1:7	8.1x10 ⁶	0	1.5x10 ⁷	12
6	1:15	4.1x10 ⁶	0	7.4x10 ⁶	4
7	1:31	2x10 ⁶	0	3.7x10 ⁶	0
8	1:63	1x10 ⁶	0	1.9x10 ⁶	0
9	1:127	5.1x10 ⁵	0	9.3x10 ⁵	0

Table 6.1 Correlation of cell viability and transformation efficiency

Electroporated samples were resuspended in 10 ml of diluent TPGY broth supplemented with 25 mM MgCl₂ and serially diluted 1:100 four times using aliquots of 100 µl. From these cultures, a further 100 µl of broth was plated on non-selective TPGY agar from which the original number of colonies were extrapolated (CFU/ml).

The results indicate that the number of CFU's is lower than 4.8×10^8 by this procedure and variable between experiments. It was evident that there was no dramatic increase in the number of transformants as the total number of cells increased. Indeed, in proportion to the number of cells present, the efficiency was reduced at higher cell concentrations. Overall the efficiency was still lower than that reported by Zhou & Johnson (1993) with the equivalent CFU's.

6.2.5.3 Effect of field strength on transformation efficiency in *C. botulinum* NCTC 2916

The electroporation method described in section 6.2.1 was repeated using the same method but the transformation was done at a variety of field strengths (kV/cm).

Sample	kV	Field Strength (kV/cm)	Number of transformants (Em ^R)	CFU/ml (3rd dilution)	% Survival (3rd dilution)	CFU/ml (4th dilution)	% Survival (4th dilution)
1	-	-	-	2.5x10 ⁹	-	3.8x10 ⁹	-
2	2.5	6.25	10	2.3x10 ⁷	0.9	0	0
3	2.25	5.625	14	3.8x10 ⁸	15.0	3.8x10 ⁸	10.0
4	2.0	5.0	11	3.3x10 ⁸	13.0	1x10 ⁹	26.0
5	1.75	4.375	23	7.3x10 ⁸	29.0	8.8x10 ⁸	23.0
6	1.5	3.75	10	7.5x10 ⁸	30.0	1x10 ⁹	26.0
7	1.25	3.125	0	1x10 ⁹	40.0	1.4x10 ⁹	36.0
8	1.0	2.5	0	8.3x10 ⁸	33.0	2x10 ⁹	53.3
9	0.75	1.875	0	1.5x10 ⁹	60.0	8.1x10 ⁹	>100
10	0.5	1.25	0	1.3x10 ⁹	50.0	6.9x10 ⁹	>100
11	0.25	0.625	0	confluent	-	1.8x10 ⁹	46.6

Table 6.2 Effect of electric field strength on transformation efficiency in *C. botulinum* NCTC 2916. Experiment 1

Electroporated samples were resuspended in 10 ml of diluent TPGY broth supplemented with 25 mM MgCl₂ and serially diluted 1:100 four times using aliquots of 100 µl. From these cultures, a further 100 µl of broth was plated on non-selective TPGY agar from which the original number of colonies were extrapolated.

Sample	kV	Field Strength (kV/cm)	Number of transformants (EmR)
1	-	-	-
2	2.5	6.25	12
3	2.25	5.625	10
4	2.0	5.0	16
5	1.75	4.375	3
6	1.5	3.75	0
7	1.25	3.125	0
8	1.0	2.5	0
9	0.75	1.875	0
10	0.5	1.25	0
11	0.25	0.625	0

Table 6.3 Effect of electric field strength on transformation efficiency in *C. botulinum* NCTC 2916. Experiment 2

The results suggest that electroporation in *G. botulinum* NCTC 2916 is not successful with an electric field strength below approximately 3.0 kV/cm. A significant increase in transformation frequency was not detected at higher field strengths.

6.2.5.4 Effect of DNA concentration on the transformation efficiency of *C. botulinum* NCTC 2916

The electroporation method described in section 6.2.1 was used with the addition of varying amounts plasmid DNA. The plasmid used was pGK12 (150 µg/ml in 1xTE buffer) prepared from *E. coli* HB101 by the CsCl method (section 2.2.2.3).

Sample	DNA (volume)	DNA (µg)	Number of transformants (EmR)
1	40.0 µl	6.000	1
2	10.0 µl	1.500	1
3	6.0 µl	0.900	2
4	3.0 µl	0.450	0
5	1.5 µl	0.225	0
6	1.0 µl	0.150	0
7	0.5 µl	0.075	0

Table 6.4 Effect of DNA concentration on transformation efficiency in *C. botulinum* NCTC 2916

The number of erythromycin resistant colonies was particularly low in this experiment which may therefore be difficult to interpret. The results also reflect the inconsistent nature and poor gene transfer efficiency experienced with this strain and with these methods.

6.2.5.5 Other parameters

Stationary growth phase ($O.D_{600} = 1.2$) and early exponential growth phase ($O.D_{600} = 0.4$) cells were electroporated under the conditions defined in section 6.2.1. Neither of these modifications increased the total number of transformants above 10. Likewise, the adjustment of resistance from between 100-800 Ω failed to improve gene transfer efficiency.

6.2.5.6 Electrotransformation of *C. botulinum* NCTC 2916 with pMTL540E

The plasmid pGK12 had indisputably been introduced into *C. botulinum* NCTC 2916, albeit at low efficiency. As the procedure clearly allowed plasmid DNA to enter the cells, the possibility existed that pMTL540E could be similarly introduced. However, during the course of over 10 attempts, not a single positive transformant was detected. To investigate this observation, alternative constructs based on the pCB102 replicon were made. The inducible MLS resistance gene in the pGK12 construct (derived from pE194; Kok *et al.*, 1984) was cloned as a PCR fragment and used to replace the constitutive MLS resistance gene present in pMTL540E (Brehm *et al.*, 1987). The β -lactamase gene from pMTL540E was also removed. The resultant plasmid functioned in *C. beijerinckii* NCIMB 8052, but again, no erythromycin resistant colonies were detected in *C. botulinum* NCTC 2916. Failure to transform could be due to the fact that, that the pCB102 replicon did not function in this group of clostridia, which is somewhat phylogenetically distinct from its natural host *C. butyricum* (Collins *et al.*, 1994), or a barrier to transformation associated with the DNA existed.

The presence of a type II restriction endonuclease in *C. botulinum* NCTC 2916 was assessed by the method of Mermelstein, (1992). Plasmids pGK12 and pMTL540E,

which were isolated from *E. coli* HB101 and used in electroporation experiments, were used as substrates in combination with commercially available restriction endonuclease buffers from Boehringer Mannheim: A, B, L, M, H and Northumbria Biochemicals Ltd: 1-13). No endonuclease activity was detected after a 16 h incubation at 37 °C. In contrast, a type II restriction endonuclease with a specificity analogous to *MspI* was detected in a non-proteolytic *C. botulinum* type B strain (T. Davis, unpublished data). Although RNA from *C. botulinum* NCTC 2916 was released into the 'cleared lysate' during the restriction endonuclease extraction procedure, light microscopy suggested that the method of Mermelstein did not efficiently lyse the group I *C. botulinum* cells. To avoid entering a protracted search for barriers which inhibited the use of pMTL540E in *C. botulinum* NCTC 2916, an alternative 'shotgun' approach was employed in which other *C. botulinum* type A strains were speculatively challenged with pMTL540E using the electroporation technique.

6.3 Electrotransformation of *C. botulinum* ATCC 3502

6.3.1 Electrotransformation

This type A strain was also described as a 'Hall strain'. Colonies were whitish-grey in appearance, mucoid and shiny. Although surrounded by 'feathery' growth, the lateral growth was not as pronounced in comparison to strain NCTC 2916.

Electroporation was conducted according to the details in section 6.2.1. pMTL540E plasmid DNA was isolated from *E. coli* HB101 by the CsCl density-gradient method (section 2.2.2.3) of which 4 µg were used in each cuvette. Using this methodology a total of 37 Em^R resistant colonies appeared after approximately 48 h incubation at 37 °C in an anaerobic cabinet.

Sample	Plasmid	Field strength	Erythromycin Resistant Colonies
1	-	6.25 kV/cm	0
2	pMTL540E	6.25 kV/cm	8
3	pMTL540E	6.25 kV/cm	17
4	pMTL540E	6.25 kV/cm	12

Table 6.5 Electroporation of *C. botulinum* ATCC 3502 with pMTL540E

6.3.2 Analysis of the putative transformants

Em^R resistant colonies were essentially morphologically identical to *C. botulinum* ATCC 3502 grown on TPGY agar. Light microscopy of the Gram stained colony preparations demonstrated that the putative transformants were short Gram positive rods. The toxigenic potential of the transformants was confirmed by amplification of a 600 bp DNA fragment of the type A neurotoxin gene (section 6.2.2). Plasmid DNA was extracted from 10 ml TPGY cultures, supplemented with erythromycin (20 µg/ml) which were grown for 16 h (section 6.2.3). *E. coli* strain TG1 was transformed with the plasmid DNA which conferred both an ampicillin and erythromycin resistant phenotype. In addition the colonies were blue in the presence of IPTG and X-GAL indicating that the *lacZα* gene was also functional. Restriction endonuclease analysis of plasmid DNA isolated from the *E. coli* transformants demonstrated that the restriction endonuclease profiles of plasmids isolated from *C. botulinum* ATCC 3502 were identical to the original vector pMTL540E.

6.3.3 Large scale plasmid isolation from *C. botulinum* ATCC 3502 transformants

Larger quantities of plasmid DNA were isolated from the *C. botulinum* ATCC 3502 transformants using the method developed for *C. botulinum* NCTC 2916 (section 6.2.4). Electrophoretic analysis of the plasmid sample in a 0.8 % (w/v) agarose gel demonstrated the presence of two high molecular weight bands. Restriction analysis revealed that the entire sample was composed of DNA fragments which corresponded to the restriction profile of pMTL540E. No visible evidence supported the existence of additional plasmids in this sample (figure 6.2).

6.3.4 Electrotransformation of *C. botulinum* ATCC 3502 with passaged DNA

To determine if the modification of plasmid DNA was influential in the frequency with which *C. botulinum* ATCC 3502 could be transformed, two sources of the plasmid were used. Preparations were made either from an *E. coli* strain (HB101 or GM2163) prepared using the CsCl density gradient method (section 2.2.2.3) or from the recipient clostridial strain, *C. botulinum* ATCC 3502 (section 6.3.3) or *C. botulinum* NCTC 2916 (section 6.2.4). The electroporation procedure was identical to that described earlier (section 6.2.1).

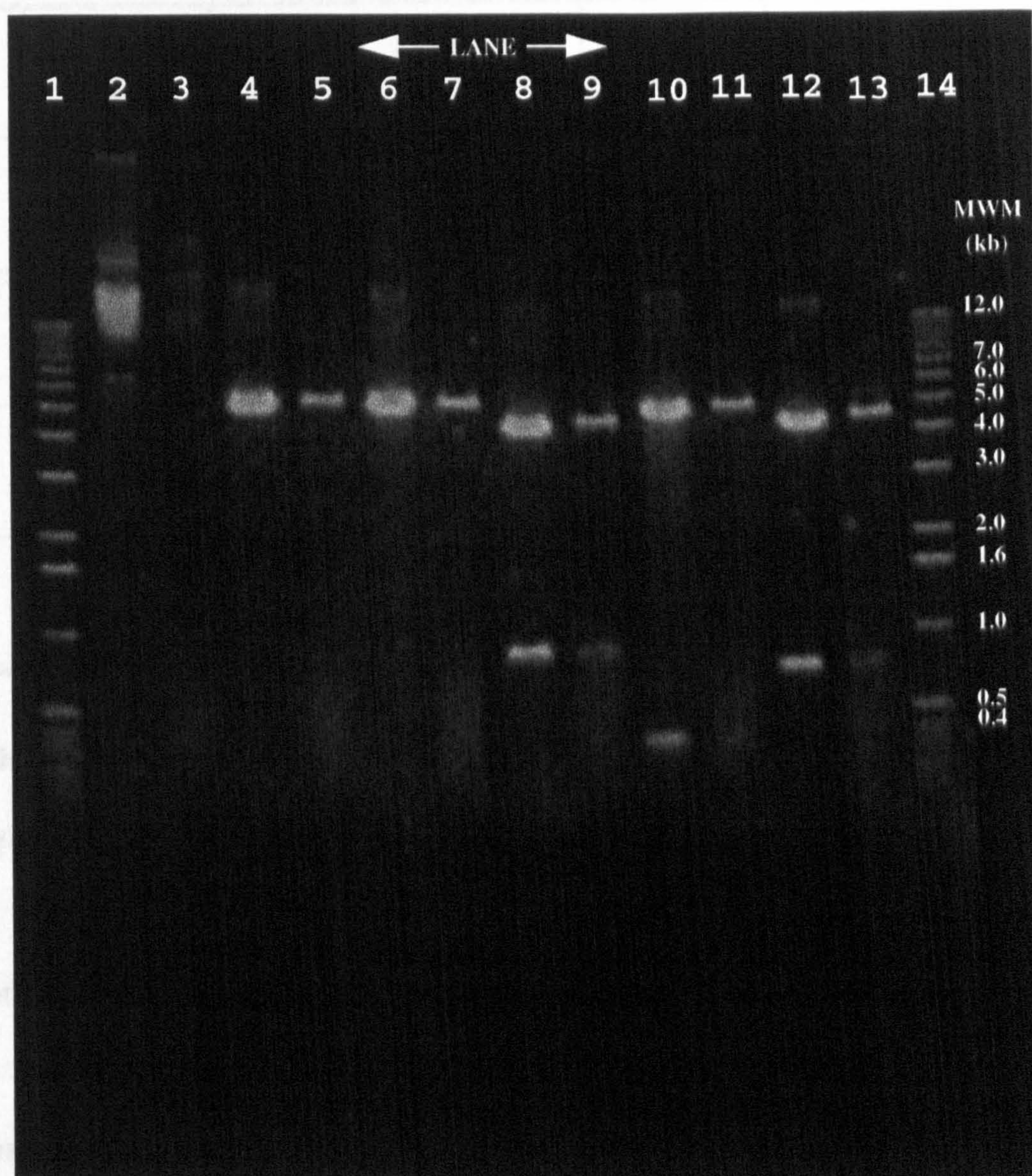


Figure 6.2 Restriction endonuclease mapping of plasmid DNA isolated from *C. botulinum* ATCC 3502

Plasmid DNA isolated from an erythromycin resistant, putative transformant colony of *C. botulinum* ATCC 3502 (section 6.3.3) was digested with a number of restriction endonucleases. pMTL540E plasmid DNA (control) isolated from an *E. coli* strain was also digested and concomitantly the samples were separated in a 0.8 % (w/v) agarose gel.

Lane 1) 1Kb DNA ladder, Lane 2) control: uncut 3) 3502: uncut 4) control: *Xba* I 5) 3502: *Xba* I 6) control: *Hind* III 7) 3502: *Hind* III 8) control: *Pvu* II 9) 3502: *Pvu* II 10) control: *EcoR* I 11) 3502: *EcoR* I 12) control: *Sca* I 13) 3502: *Sca* I 14) 1Kb DNA ladder.

The DNA fragments demonstrate that the plasmid DNA isolated from *C. botulinum* ATCC 3502 is essentially indistinguishable from pMTL540E isolated from *E. coli* indicating that this clostridial strain is successfully transformed with plasmid.

Plasmid	Source	Recipient	Experiment 1	Experiment 2	Experiment 3
pMTL540E	<i>E. coli</i> (HB101)	<i>C. botulinum</i> ATCC 3502	5	7	8
pMTL540E	<i>C. botulinum</i> ATCC 3502	<i>C. botulinum</i> ATCC 3502	6	11	3
pGK12	<i>E. coli</i> (GM2163)	<i>C. botulinum</i> ATCC 3502	0	3	6
pGK12	<i>C. botulinum</i> NCTC 2916	<i>C. botulinum</i> ATCC 3502	0	0	0

Table 6.6 Electroporation of *C. botulinum* ATCC 3502 with plasmid DNA from different sources

The transformation results did not provide sufficient evidence to speculate whether DNA modification was an important factor in determining transformation efficiency. Each plasmid preparation was able to transform *E. coli* strains at high efficiency indicating that the DNA was of good quality.

6.3.5 Effect of a cell wall modification agent on transformation efficiency in *C. botulinum* ATCC 3502

The effect of glycine on transformation efficiency was established by first determining the concentration at which it exerted an inhibition of cell growth. Incremental concentrations of glycine (w/v) in TPGY broth were inoculated with a 3 % volume of fresh exponential culture of *C. botulinum* ATCC 3502 and incubated at 37 °C. After 16 h broths containing glycine up to a 2 % (w/v) concentration appeared turbid, whilst

those at 2.5 % (w/v) and 3.0 % (w/v) appeared weakly turbid after 48 h. A repetition of this experiment with 0.1 % (w/v) increments between 2.0 % (w/v) and 3.0 % (w/v) glycine indicated that 2.2 % - 2.3 % (w/v) glycine concentration range was critical for growth inhibition. Electrotransformation was repeated as detailed in section 6.2.1 except that *C. botulinum* ATCC 3502 was cultured overnight in 30 ml TPGY broth supplemented with 2 % glycine (w/v) and inoculated into 2 x 300 ml TPGY broth supplemented with 2 % (w/v) glycine for the preparation of electro-competent cells.

Plasmid	Source	Recipient	Experiment 1	Experiment 2	Experiment 3
pMTL540E	<i>E. coli</i> (HB101)	<i>C. botulinum</i> ATCC 3502	276	49	72
pMTL540E	<i>C. botulinum</i> ATCC 3502	<i>C. botulinum</i> ATCC 3502	6	12	ND
pGK12	<i>E. coli</i> (GM2163)	<i>C. botulinum</i> ATCC 3502	9	23	10
pGK12	<i>C. botulinum</i> NCTC 2916	<i>C. botulinum</i> ATCC 3502	0	ND	ND

Table 6.7 Effect of a cell wall modification agent on gene transfer frequency in *C. botulinum* ATCC 3502

The results indicate that the addition of sub-inhibitory concentrations of glycine are able to increase electrotransformation efficiency by a modest amount. When comparing DNA isolated from *E. coli* HB101 this increase approached a one log increase. However the success of this procedure still remains inconsistent.

6.4 Characterisation of a plasmid from *C. botulinum* NCTC 2916

6.4.1 Justification for cloning a plasmid indigenous to a group I, *C. botulinum* strain

During the course of the gene transfer studies in *C. botulinum* NCTC 2916, the failure to obtain a transformant using pMTL540E suggested that the replicon of this plasmid was unsuitable for use in group I *C. botulinum*. The possibility of isolating an alternative replicon derived from this organism was perceived as a means of circumventing such a problem. The co-isolation of a putative indigenous plasmid with pGK12 facilitated this process (section 6.2.4). The rational characterisation of the plasmid isolated from *C. botulinum* NCTC 2916 was implemented.

6.4.2 Restriction endonuclease mapping of the putative plasmid from *C. botulinum* NCTC 2916

Samples of plasmid DNA isolated from *C. botulinum* NCTC 2916 (section 6.2.4) were digested with a range of restriction endonucleases and analysed on agarose gels. These gels indicated that an extrachromosomal element, probably a plasmid of approximately 6.0 kb, appeared to be linearised with either *EcoRI*, *XbaI* or *BglII* without cutting pGK12. Another enzyme, *SacI*, could linearise both pGK12 (4.4 kb) and the indigenous plasmid leaving a faint band which was refractory to digestion by any enzyme used. It was thought that this band consisted of single stranded DNA plasmid intermediates (figure 6.1). Digests with *HindIII* appeared to cut the indigenous plasmid into two fragments of approximately 3.1 kb and 2.7 kb, whilst a second enzyme, *HpaII*, produced bands of approximately 4.2 and 1.5 kb.

6.4.3 Cloning a cryptic plasmid from *C. botulinum* NCTC 2916

The two *Hind*III fragments from the cryptic plasmid appeared to be the most suitable candidates for cloning, partly because the plasmid could be sequenced from four separate ends and the original plasmid could be easily reconstructed. An 80 µl sample of the plasmid preparation from *C. botulinum* NCTC 2916 (section 6.2.4) was digested with *Hind*III and purified by phenol extraction. Likewise, 10 µg of the *E. coli* cloning vector pMTL20 was digested with *Hind*III and phenol extracted. The two preparations were ligated together and used to transform *E. coli* DH5 α. Transformants were selected in the presence of IPTG and X-Gal. White colonies were selected and screened for plasmids containing inserts of approximately 3.1 kb or 2.7 kb. Two colonies were found to contain inserts of the expected size. The plasmids pBP19 and pBP2, were used for further analyses.

6.4.4 Determination of the complete nucleotide sequence of the indigenous plasmid from *C. botulinum* NCTC 2916

6.4.4.1 Nucleotide sequence analysis

Plasmid DNA of the clones pBP19 and pBP2 were purified using the Qiagen method. Approximately 1 µg of each plasmid was sequenced with universal or reverse universal oligonucleotide primers (Appendix V) using an ABI automated sequencer (section 2.2.11.5). The entire nucleotide sequence of each clone was determined on both strands using an oligonucleotide primer 'walking' method. To establish the orientation of the two DNA fragments and contiguity of the *Hind*III sites, four sequencing primers were used as PCR primers to amplify DNA fragments spanning both *Hind*III sites. The PCR products were cloned (section 2.2.12.4) and sequenced. Individual sequences were

assembled into a single contig using the DNASTAR software package 'Seqman'. The contig was then translated in all six reading frames using the DNASTAR software package 'Mapdraw' (figure 6.3). Assembly of the DNA sequences yielded the putative *C. botulinum* plasmid pBP1, of 5927 nucleotides in length and with an A+T base composition of 74.34 % typical of mesophilic clostridial DNA.

6.4.4.2 Analysis of putative ORFs

An analysis of the translated DNA sequence suggested that pBP1 consisted of 8 ORFs which were encoded on either DNA strand (see table 6.9 and figure 6.4). Distinct operons appeared to be formed by ORFs 3 and 4 plus ORFs 6 and 7. ORF 2 assumed a divergent orientation to ORF 3, whilst a similar relationship was found between ORF 5 and ORF 6 (figure 6.4). To assign functions to the putative proteins encoded by ORFs 1-8, each translated protein sequence was used to search the SwissProt data base using a low K-tuple and percentage threshold value. In addition, the secondary structure for each putative protein was predicted using the DNASTAR program 'Protean'.

ORF 1

This ORF does not possess an obvious ribosome binding site or translational initiation codon. Comparison of the entire ORF with the current DNASTAR EMBL database (issue 100) suggested that the product of this gene bears a high degree of similarity with a group of topoisomerases with replicative origin-specific activities. This class of proteins are implicated in initiating the replication of plasmids whose members are broadly represented by pC194, a plasmid which replicates via a RC mechanism (Novick, 1989). An alignment of several replication proteins most similar to ORF 1 (figure 6.5)

Hind III
AAGCTTCAAATAATATTGATGAATATGCTGAAAAATTAAAAGTTAATTTATCATCTAATTCTAATGAAATGCCAATATCAAGGGCATATTCAGGTTACAA
----- 100
TTCGAAGTTTATTATAACTACTTATACGACTTTTAAATTTCAATTAAATAGTAGATTAAAGATTACTTTACGGTTATAGTTCCCGTATAAGTCCAATGTT
A S N N I D E Y A E K L K V N L S S N S N E M P I S R A Y S G Y N

TAGTTCGGTGCTGTAGCTTATGCCATAAATGGGCTACCTCTTTTAAATCCTAAGTATATGAGATTTTCTGCTGATTGTACAAATTTGTATCTCAATGC
----- 200
ATCAAGACCACGACATCGAATACGGGTATTTACCCGATGGAGAAAATTAGGATTCAATACTCTAAAAGACGACTAACATGTTTAAACATAGAGTTACG
S S G A V A Y A H K W A T S F N P K Y M R F S A D C T N F V S Q C

EcoR I Sac I
GCTTTTGAAGGTGGAAGATTCCAAACCATAATATTTACCAGTTTGGATACCAGCACTACAAGAAAAAGGAGCTCCAAATATTCAAATCATGGACTT
----- 300
CGAAACTTCCACCTTCTTAAGGTTTGGTATTATAAAGTGGTCAAACCTATGGTCGTGATGTTCTTTTCTCGAGGTTTATAAGTTTGTACCTGAA
A F E G G R I P N H N I S P V W I P A L Q E K G A P K Y S K S W T C

GCGTAAAAGATTTTATACTTTATGCTTCAAGCAGGCTATGCTTCAGCTCCAGAAGGCGGTTGTAAACAAGCTAGATTGGGTGATGTTGCAGCTTTAT
----- 400
CGCATTTCTAAAAATATTGAAATACGAAGTTCGTCCGATACGAAGTCGAGGTCTTCCGCCAACATTTGTTGATCTAACCCACTACAAACGTGAAATA
V K D F Y N F M L Q A G Y A S A P E G G C K Q A R L G D V C S F I

AATGTACCTAAAAATGATTGGACTCATGCAATGATTGTAACATAAATGATAGTAAAGGACGTGTGTATTATTGAGCACATACTGACCCTAAATGGATG
----- 500
TTACATGGATTTTACTAACCTGAGTACGTTACTAACATTGATTTTAACTATCATTTCCTGCACACATAATAAGTCGTGTATGACTGGGATTTAACCTAC
M Y L K M I G L M Q M . I V T K I D S K G R V Y Y S A H T D P K L D V
orf 4 ➡

TTCCTGTTTGGTACGCTTTAGATGGTAAAACCTTATGGGAACATAAGATTAATAAAATTTTGGCATTAAAATAATTTAAATTACTCTTGTACTTAAAAGAA
----- 600
AAGGACAAACCATGCGAAATCTACCATTTTGAATACCCCTGTATTCTAATTATTTTAAAACCGTAATTTTATTAAATTTAATGAGAACATGAATTTTCTT
P V W Y A L D G K T Y G N I R L I K F W H .

(Transcriptional terminator) Bgl II
AGGAACTTGTAAAAGTTCCTTTCTTTTATTCTTTATTATAAAGTTAAGTTTAAATTTATAATATGTTTTCATCTAAATGAGTTAAATGTTAAAAACGA
----- 700
TCCTTGAACATTTTCAAGGAAAGAAAATAAGAAATAATATTTCAATTCAAATTTTAAATATTATACAAAATCTAGATTTACTCAATTTACAATTTTTTGCT

Bgl II
AAAATAATATAGGTTTAAATAAATAGTCATTAAATAGCTTTTTTATTATATAAAAAGTTTATAGCAAGATCTTCAAGAATTTGAGATGCGGTTGTAT
----- 800
TTTTATTATATCCAAATATTTTAAATCAGTAATTTTATCGAAAAAATAAATATATTTTCAAATATCGTTCTAGAAGTTCTTAAAGTCTACGCCAACATA
. K Y L F N I A L D E L I E S A T T

Taq I
GTTCTTTTATACTTTTTATTCTTAAATCCGTCCAAACATCATCATCGAAAATTACTGTATGCTTACTGCTCTTATTGTTTTTTTTGGGTCTGCCGACTTT
----- 900
CAAGAAAATATGAAAAATAAGAATTTAGGCAGGTTTGTAGTAGTAGCTTTTAAATGACATACGAATGACGAGAATAACAAAAAAACCCAGACGGCTGAAA
H E K I S K I R L D T W V D D D F I V T H K S S K N N K K P R G V K

TATTTCATCTTTACATATTTCTTTAGGCTCTATGGTTTTAAATTCATTATTTTTTAGTATATATCCATATCTTTTAACTTAGATGTTATGCCCTTGAT
----- 1000
ATAAAGTAGAAATGTATAAAGAAATCCGAGATACCAAATTTAAGTAATAAAAAATCATATATAGGTATAGAAAATTTGAATCTACAATACGGGGAACATA
I E D K C I E K P E I T K F E N N K L I Y G Y R K F K S T I G R S

ATGCCTAAATTTTGTAGATATTTTGTCTATTGATTTCCTCGTTCATCAATTGAGAATTAAAAACTCTATCTGCTTATCAATAGTTAAATTTAAAAATCTT
----- · 1100
TACGGATTTAAAAATCTATAAACGAATAACTAAAGGGCAAGTAGTTAAGTCTTAATTTTGTAGATAGACGAATAGTTATCAATTTAAATTTTAAGAA
I G L N K S I K S I S K G N M L E S N F F E I Q K D I T L N L F E

CTCTATCCATACATAAACCTCCTATATAGTTTAAATCTGCATATATTATATTAACATATTTACAAAAATATGGAAATTAAATATTAGTATATGTCTAT
----- · 1200
GAGATAGGTATGTATTTTGGAGGATATATCAAAATTTAGACGTATATAATATAATTGTATAAATGTTTTATACCTTTAATTTATAATCATATACAGATA
E R D M orf5 RBS

CTTTTTTATGTTTAGTAGTTATAGAAATCCTAAATTAATGGCACATATAAGAAGTAAATTTTACAAAGTTGCCTTATTTTATATATAAAAAAATTTT
----- · 1300
GAAAAAATACAAATCATCAATATCTTTAGGATTAAATTACCGTGTATATTCTTCATTTAAATGTTTCAACGGAATAAAATATATATTTTTTTTTAAAAA

TTTTATTCTTGTTATTTTTTTTTTAATAACTTGTATATAAGAAAAAACAACAAAAAATAAGGAGGAAAACATTATGAAGAAAAATAAGAAAAACAA
----- · 1400
AAAATAAAGAACAATAAAAAAAATTTATGAACATATATTCTTTTTTTGTTTTTTTTATTCCTCCTTTTGATAATACTTCTTTTATTCTTTTTGTTT
RBS orf6 M K K N K K N K

TTAAAGAAAAATTTATTAAGAAGATAGAAATTTTACCTGTTGCACCTTTTTATTAATTGTGTTTTTAATATTTTAAAAAATTTTAAAGATATATTT
----- · 1500
AATTTCTTTTAAATAATTTCTTCTATCTTTAAAAAATGGACAACGTGAAAAATAATTAACACAAAAATTATAAAATTTTAAATTTCTATATAAAA
L K K N L L K K I E I F L P V A L F I N C V F N I L K N F K D I F F orf7

TTAATGAAGTATATATTATACTTTATAATTTTATGCTCTGTTTAACTCTATCTATGCTCTATTTTACTCTGTTTAACTCCATTTTACTCTATTA
----- · 1600
AAATTACTTCATATATAATATGAAATATTAAAAATACGAGACAAAATTGAGATAGATACGAGATAAAATGAGACAAAATTGAGGTAAAAATGAGATAAT
. M K Y I L Y F I I F M L C F N S I Y A L F L L C F N S I F T L L

TTTTTGATTGAAATGGAGTCAATGTGGAGATTTTAAAAATTTTATGTAGGTACGGAATAAAAAATTTTAAATACTACATTGATGGAATGGAAATCATC
----- · 1700
AAAACTAACTTTACCTCAGTTACACCTCTAAAATTTTAAAAATACATCCATGCCTTATTTTTTAAATTTTATGATGTAACCTACCTTACCTTTAGTAG
F L I E M E S M W R F .

ACACTGGAAGCCCAAGGGGGCAACATACAATGTTGCTCCCGCGCCTAACGGCAAGTGATTATGGGTAAAGGGGTACCCCTTTTGTCTGTCGGGCAACC
----- · 1800
TGTGACCTCGGGGTTCCCGCGTTGTATGTTACAACGAGGGGCGCGGATTGCCGTTCACTAATACCCAATTCGCCCATGGGGAAAACGACAGCCCGTTGG

CCGACCCCTCGGCAGAGCATAGGACGGAGTCTATAAAGGGCAACGCCCTTAAAGCCTCGCAGAGCATATTTACTTTAGGCATAAAGTATAATATGCTA
----- · 1900
GGCTGGGGAGCCGTCTCGTATCCTGCCTCAGGATATTCCCGTTGCGGGAAATTCGGAGCGTCTCGTATAAATGAAATCCGTATTTTCATATTATACGATA

-10
ACTTTACTTAAAGTATTTAAAAAAGTGTAGTAGTAAAAAGGAGTTTGTATGTGAGTTATTTAGTTTGCACAATGAAGAAATGAAAAAGATAATTTA
----- · 2000
TGAAATGAATTTTCATAAATTTTGTGACATCATTTTTTCTCAAACTACACTCAATAAATCAAACGTGTTACTTCTTTACTTTTTCTATTAAAT
(Pre recombinase) RBS orf8 V S Y L V C T M K K M K K D N L

TTTGGTTTGCAAAACATAATCAACGAGAATCTGAAAATCATAATAACAAAGATATTGATAAAAAATAAAAGTGATTTGAATTATGATTTAGTAAATGAAA
----- 2100
AAACCAAACGTTTTTGTATTAGTTGCTCTTAGACTTTTAGTATTATTGTTTCTATAACTATTTTTATTTTCACTAAACTTAATACTAAATCATTTACTTT
F G L Q K H N Q R E S E N H N N K D I D K N K S D L N Y D L V N E N

ATAAAATTAATCTCTCAGAAAAGGTATTTGATTTGATTGAAAAATAAGAGAAAGTGAAAAAGCTATCCGTAAGGATGCTGTTTTACTAGATGAATTTAT
----- 2200
TATTTTAATTAGAGAGTCTTTTCCATAAACTAACTAACTTTTTATTCTCTTTCACTTTTTCGATAGGCATTCTACGACAAAATGATCTACTTAAATA
K I N L S E K V F D L I E K I R E S E K A I R K D A V L L D E F I

TATAAGCAGCGATAGAGATTTTTTTAAAAATATGAATGAAAATGAAATTA AAAAGTTCTTTGAAACTTCAAAACAGTTTTTTGCAGAGCGTTATGGGGAA
----- 2300
ATATTCGTCGCTATCTCTAAAAAAATTTTTATACTTACTTTTACTTTAATTTTTCAAGAACTTTGAAGTTTTGTCAAAAAACGTCTCGCAATACCCCTT
I S S D R D F F K N M N E N E I K K F F E T S K Q F F A E R Y G E

Nde I
GAAAATATAATTAGTGCTAATGTCCATTTAGATGAAACTACCCCTCATATGCACCTAGATGTGTGTCCGTTACGCAATGGGAGACTGCAAAGTAAAAATA
----- 2400
CTTTTATATTAATCACGATTACAGGTAAATCTACTTTGATGGGGAGTATACGTGGATCTACACACAGGCAATGCGTTACCCCTGACGTTTCATTTTTAT
E N I I S A N V H L D E T T P H M H L D V C P L R N G R L Q S K N I

TTTTTACTAGAAATGAACTTAGAAATATTCAAGAAGAATTACCTAAAACTTTAAAAATAATGGATTTGAAATTGAGAGAGGTGTTAAAGGATCTACTCA
----- 2500
AAAAATGATCTTTACTTGAATCTTTATAAGTTCTTCTTAATGGATTTTGAAATTTTTATTACCTAAACTTTAACTCTCTCCACAATTTCCTAGATGAGT
F T R N E L R N I Q E E L P K T L K N N G F E I E R G V K G S T Q

AAAACATATAGACATACAACTTTAAAAGCACGAGAAGAAAAAACTTAAAGAGTTAAAGGAAAAAATAAAGAGATAGAATTAAAGATTGCCAGTAGA
----- 2600
TTTTGTATATCTGTATGTTTGAAATTTTCGTGCTCTTCTTTTTTTGAATTTCTCAATTCCTTTTTTTATTTCTCTATCTTAATTTCTAACGGTCATCT
K H I D I Q T L K A R E E K K L K E L K E K N K E I E L K I A S R

AAAGAAAAATTAGAAGAAGTTAAAGATTTAGATTTTCAAATTAAATCAAAAAAACTGTTTTAAATAACTTAAATCAAGGGAATGACTGTCTCGATGAAG
----- 2700
TTTCTTTTAAATCTTCTTCAATTTCTAAATCTAAAAGTTTAAATTTAGTTTTTTTTGACAAAAATTATTGAATTTAGTTCCTTACTGACAGAGCTACTTC
R E K L E E V K D L D F Q I K S K K T V L N N L N Q G N D C L D E D

Hind III
ATATTAAATATAAGCTTGATGGTTTTGACGTCAAAATGGATTACGTTGAATTTAGAAAATTTCAAACCTCAAATTGCAAGTATTAAAAAAGTAAAAAAGA
----- 2800
TATAATTTATATTGAACTACCAAACTGCAGTTTTACCTAATGCAACTTAAATCTTTTAAAGTTTGAGTTTAAAGTTTCAATAATTTTTTCATTTTTTCT
I K Y K L D G F D V K M D Y V E F R K F Q T Q I A S I K K V K K E

AAATGAAGATTTAAAAAATGAAAATAGTATAGTTCATAAAAAATGTTCTGACCTTAGAAAATCTCAAGAAGAAGTATCTAGTAGGTTTACAAATTGGATA
----- 2900
TTTACTTCTAAATTTTTTACTTTTATCATATCAAGTATTTTTTACAAGACTGGAATCTTTTAGAGTTCTTCTTATAGATCATCCAAATGTTTAACTTAT
N E D L K N E N S I V H K K C S D L R K S Q E E V S S R F T N W I

Nde I Xho I / Taq I
GATAATACTTACAAGCATATGTCTTTGGACTCGAGAGAGCAGTTCAAACGAAGAGGTTCTGAATCCTTAGCTAATGGTTCAACAGGTAACCTATGACGAAG
----- 3000
CTATTATGAATGTTTCGTATACAGAAACCTGAGCTCTCTCGTCAAGTTTGCTTCTCCAAGACTTAGGAATCGATTACCAAGTTGTCCATTGATACTGCTTC
D N T Y K H M S L D S R E Q F K R R G S E S L A N G S T G N Y D E D

ATAGCACCTGGATAAGTCTGTAATGGATTCTAAGGCATTTAATGAAGACGTGTATATAAAATGTGCTAATGAAAAAGAAAATGCGTTAAAAGAGCCTAA
----- 3100
TATCGTGGGACCTATTCAGACATTACCTAAGATTCCGTAAATTACTTCTGCACATATATTTTACACGATTACTTTTTCTTTTACGCAATTTTCTCGGATT
S T L D K S V M D S K A F N E D V Y I K C A N E K E N A L K E P K

AATGAGTTCAAATGGTTTTGAAATTGATTGGTAGTTTAAATTAATATATTTTTCTATTGGCTATCTCGATACCTATAGAATCTTCTGTTCACTTTTGTT
----- 3200
TTACTCAAGTTTACCAAACTTTAACTAACCATCAAATTAAATTATATAAAAAAGATAACCGATAGAGCTATGGATATCTTAGAAGACAAGTGAAAACAA
M S S N G F E I D W .

TTTGAAATATAAAAAGGGCTTTTAGCCCCCTTTTTTAAAACCTCCGGAGGAGTTTCTTCATTCTTGATACTATACGTAACCTATTTTCGATTGACTTC
----- 3300
AACTTTATATTTTCCCGAAAAATCGGGGAAAAAAATTTTGAGGCCTCCTCAAAGAAGTAAGAACTATGATATGCATTGATAAAAGCTAAACTGAAG

ATTGTCAATTAAGCTAGTAAAATCAATGGTTAAAAAACAAAAACTTGCATTTTCTACCTAGTAATTTATAATTTTAAAGTGTGAGTTTAAAAGTATAA
----- 3400
TAACAGTTAATTCGATCATTTTAGTTACCAATTTTTTGTTTTTGAACGTAAAAAGATGGATCATTAAATATTAAAATTCACAGCTCAAATTTTCATATT

TTTACCAGGAAAGGAGCAAGTTTTTAATAAGGAAAAATTTTTCCTTTTAAAATCTATTTTCGTTATATGACTAATTATAATCAAAAAATGAAAAATAA
----- 3500
AAATGGTCCTTTCCTCGTTCAAAAAATTATTCCTTTTAAAAGGAAAAATTTTAAGATAAAGCAATATACTGATTAATATTAGTTTTTTTACTTTTATTT

CAAGAGGTAAAAACTGCTTTAGAGAAATGTACTGATAAAAAAGAAAAATCCTAGATTTACGT CATACATAGCACCTTTAACTACTAAGAAAAATATTG
----- 3600
GTTCTCCATTTTGGACGAAATCTCTTTACATGACTATTTTTTCTTTTTTAGGATCTAAATGCAGTATGTATCGTGGAATTTGATGATTCTTTTTATAAC
(Rep) RBS orf1 → L E K C T D K K R K N P R F T S Y I A P L T T K K N I E

AAAGGACTTCCACTTGTGGAGATTATTTGTTTATGTTGAGTGATGCAGACTTAGAACATTTTAAATTACATAAAGGTAATTTTGCGGTAATAGATTTTG
----- 3700
TTTCCTGAAGGTGAACACCTCTAATAAACAAATACAACCTCACTACGTCTGAATCTTGTAATAATTTAATGTATTTCCATTAAAAACGCCATTATCTAAAC
R T S T C G D Y L F M L S D A D L E H F K L H K G N F C G N R F C

TCCAATGTGTAGTTGGCGACTTGCTTGTAAGGATAGTTTAGAAATATCTATTCTTATGGAGCATTTAAGAAAAGAAGAAAATAAAGAGTTTATATTTTA
----- 3800
AGGTTACACATCAACCGCTGAACGAACATTCCTATCAAATCTTTATAGATAAGAATACCTCGTAAATTCCTTCTTTTATTTCTCAAATATAAAAT
P M C S W R L A C K D S L E I S I L M E H L R K E E N K E F I F L

ACTCTTACAACCTCCAAATGTAAAAAGTTATGATCTTAATTATTCTATTAAACAATATAATAAATCTTTTAAAAAATTAATGGAGCGTAAGGAAGTTAAGG
----- 3900
TGAGAATGTTGAGGTTTACATTTTCAATACTAGAATTAATAAGATAATTTGTTATATTATTTAGAAAATTTTAAATTACCTCGCATTCCTTCAATTCC
T L T T P N V K S Y D L N Y S I K Q Y N K S F K K L M E R K E V K D

ATATAACTAAAGGTTATATAAGAAAATTAGAAGTAACTTACCAAAAGGAAAAATACATAACAAAGGATTTATGGAAAATAAAAAAGATTATTATCAAAA
----- 4000
TATATTGATTTCCAATATATTCTTTAATCTTCATTGAATGGTTTTCTTTTTATGTATTGTTTCCTAAATACCTTTTATTTTTTTCTAATAATAGTTTT
I T K G Y I R K L E V T Y Q K E K Y I T K D L W K I K K D Y Y Q K



Figure 6.3 Complete annotated nucleotide sequence of pBP1

Both DNA strands are shown and the *Hae* III site is duplicated at either end. Reading frames which use the upper strand as sense strand are shown in purple whilst reading frames which correspond to the lower strand are shown in blue. Possible secondary structures are represented as converging arrows and those thought to form the *45S* (minic origin) are in red. Putative ribosome binding sites (RBS) are marked in red and termination codons are represented with a single dot. Other features are labeled accordingly.

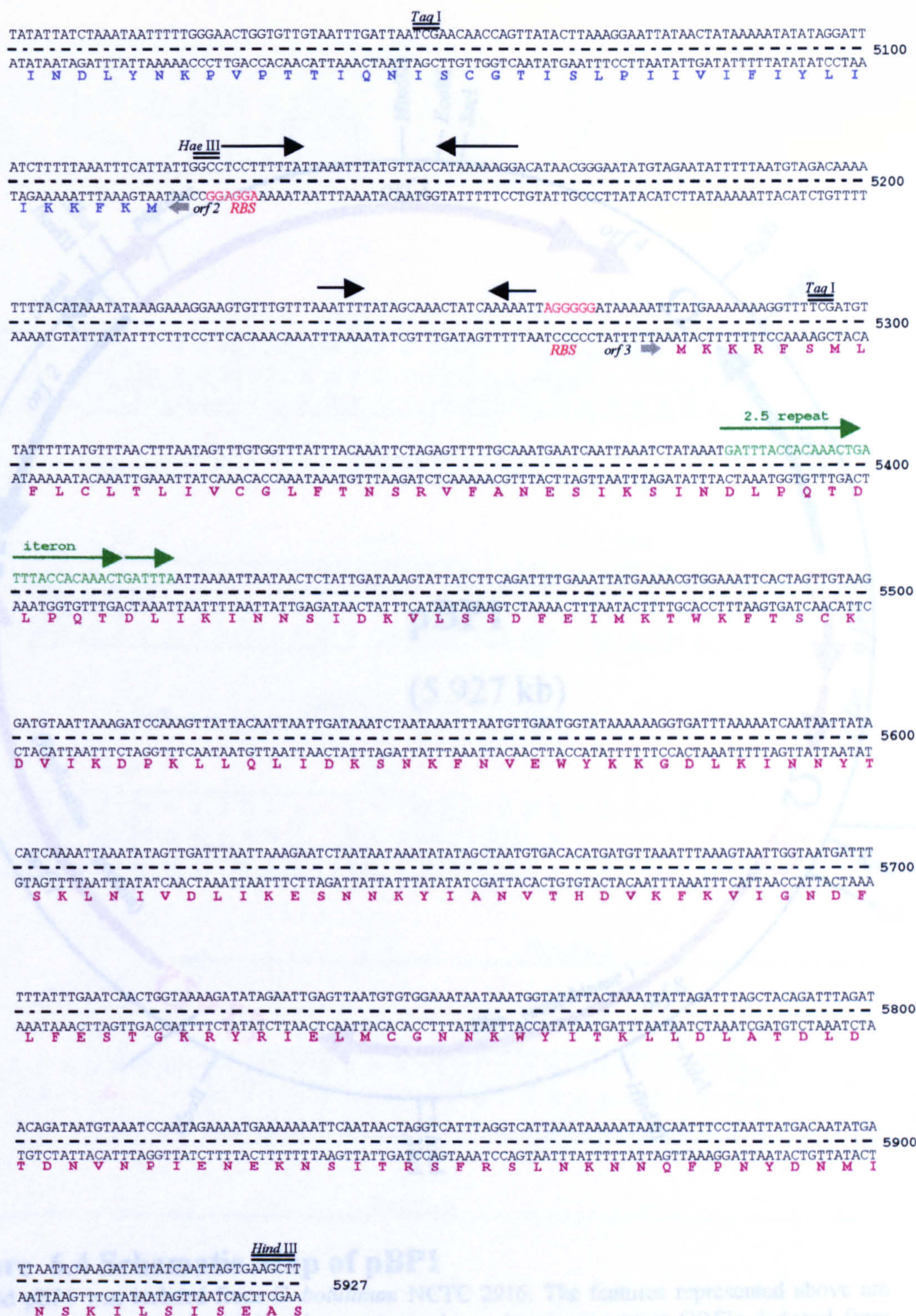


Figure 6.3 Complete annotated nucleotide sequence of pBP1

Both DNA strands are shown and the *Hind* III site is duplicated at either end. Reading frames which use the upper strand as sense strand are shown in purple whilst reading frames which correspond to the lower strand are shown in blue. Possible secondary structures are represented as converging arrows and those thought to form the *dso* (minus origin) are in tan. Putative ribosome binding sites (RBS) are shaded in red and termination codons are represented with a single dot. Other features are labeled accordingly.

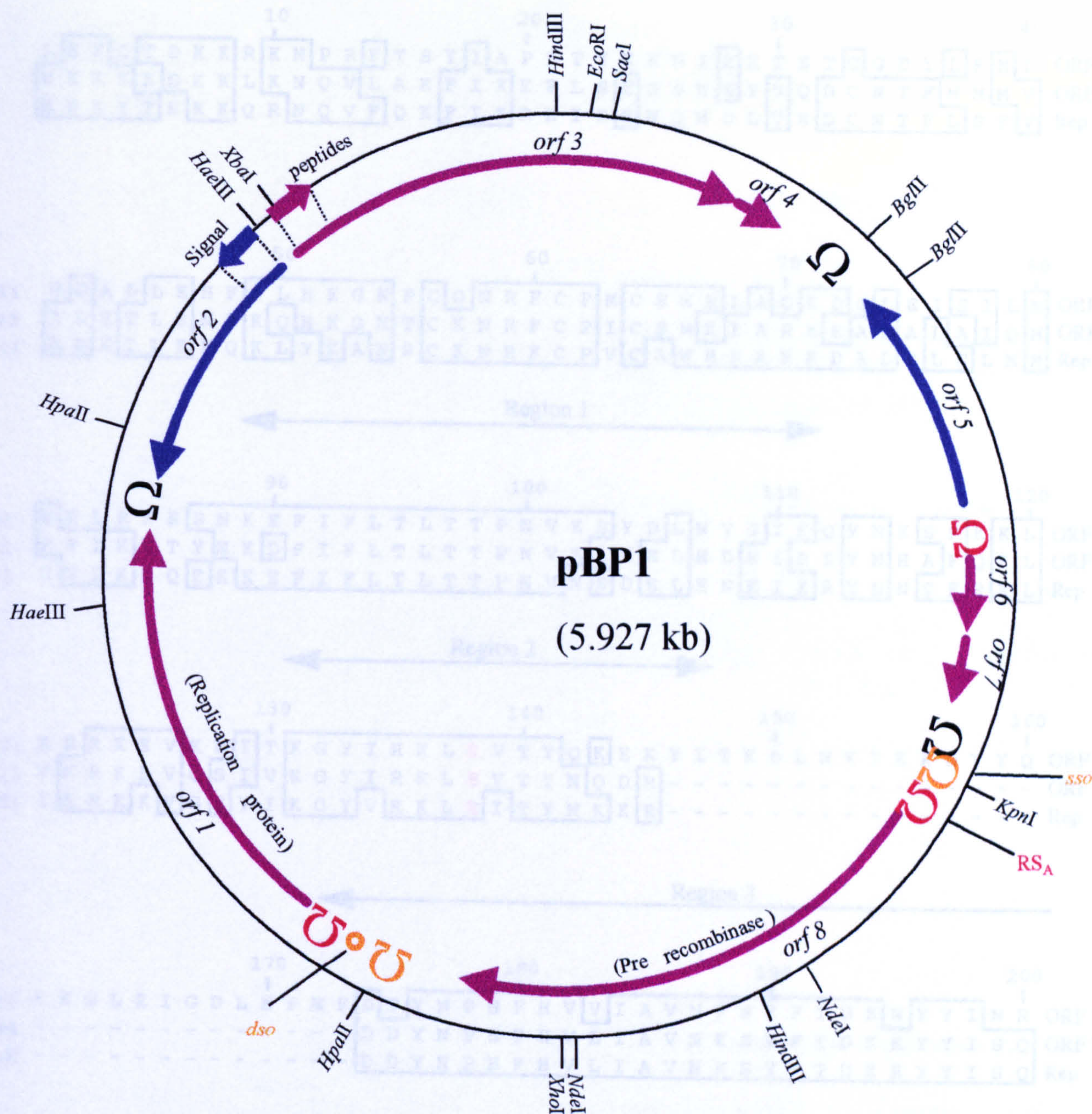


Figure 6.4 Schematic Map of pBP1

Plasmid pBP1 was isolated from *C. botulinum* NCTC 2916. The features represented above are derived from analysing the nucleotide sequence alone. Arrows represent ORF's deduced from combinations of translational signals, Ω represents dyad repeats typical of Rho-independent transcriptional terminators, Ω represent dyad repeats occurring in possible mRNA leader sequences which may function in transcriptional control, Ω alone represent dyad repeats which could form a putative minus/single stranded origin of replication (ssO). Ω and ○ together represents a sequence resembling the double stranded origin of replication from the pC194 group of rolling circle plasmids (Seery *et al.*, 1993).

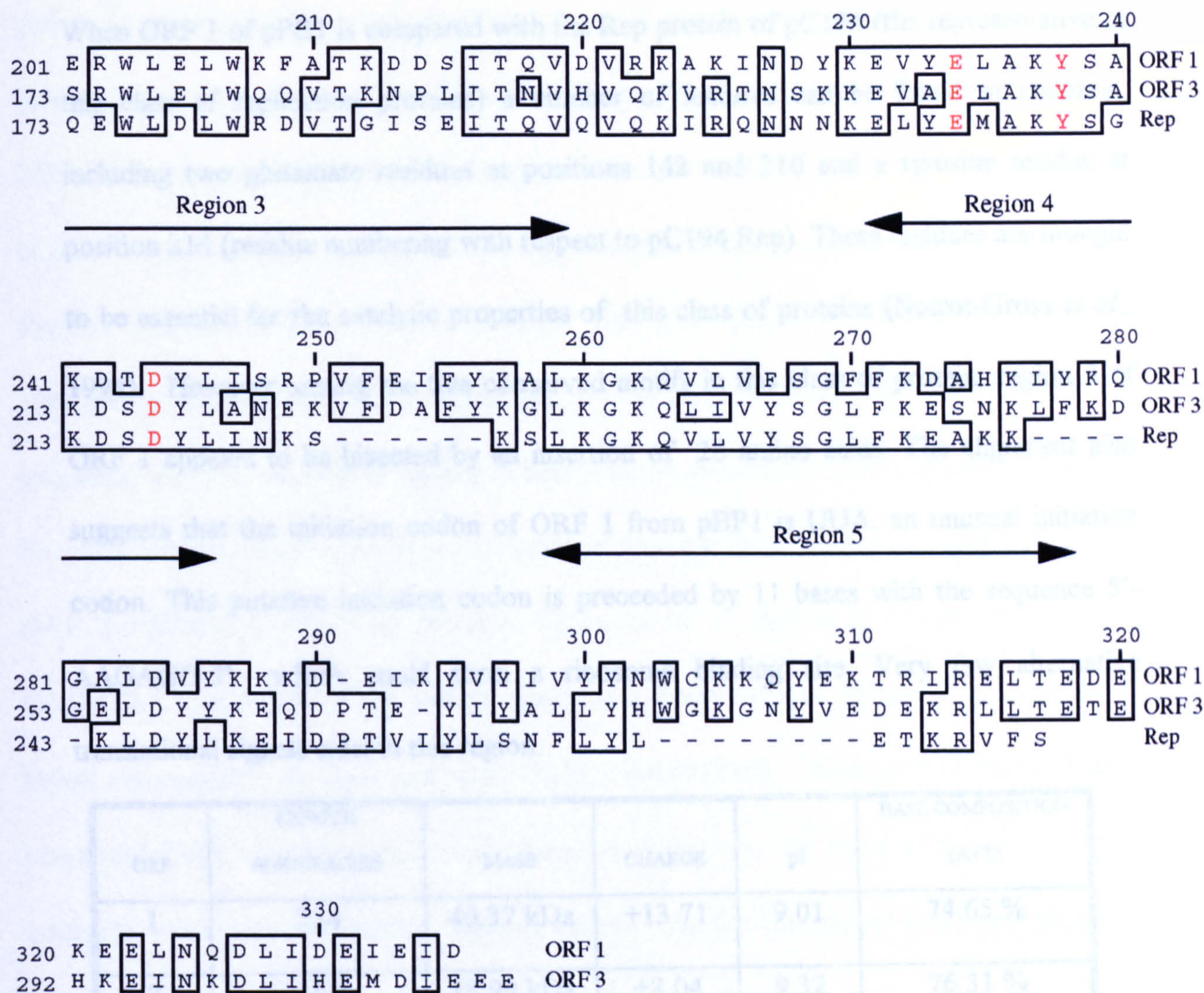


Figure 6.5 Comparative alignment of the pBP1 ORF 1 with pC194 type replication proteins

ORF 1 is the putative replication protein from pBP1 (this study), ORF 3 is a putative replication protein from the bacterium *Marinococcus halophilus* (Louis & Galinski, 1996) and Rep is the replication protein from the *Staphylococcus aureus* plasmid pC194 (Horinuchi & Weisblum, 1982) which represents this family. Amino acids are boxed when two or more residues are identical. Homologous regions of the three proteins suggest that they may all be members of the pC194 group of plasmid replication proteins. Regions I-V of the pC194 group are depicted by arrows with are beneath the sequences (adapted from Noirot-Gros *et al.*, 1994). Residues which appear to be equivalent to the catalytic residues predicted in the active site of pC194 are in red (Noirot-Gros *et al.*, 1994).

demonstrated that the primary sequence of ORF 1 essentially conforms to the five conserved regions described for this class of replication proteins (Noirot-Gros *et al.*, 1994).

When ORF 1 of pPB1 is compared with the Rep protein of pC194 (the representative of this class of replication proteins) a number of features can be found in common, including two glutamate residues at positions 142 and 210 and a tyrosine residue at position 214 (residue numbering with respect to pC194 Rep). These residues are thought to be essential for the catalytic properties of this class of proteins (Noirot-Gross *et al.*, 1994). However among the five conserved motifs in this class of protein, region 3 of ORF 1 appears to be bisected by an insertion of 28 amino acids. The alignment also suggests that the initiation codon of ORF 1 from pBP1 is UUA, an unusual initiation codon. This putative initiation codon is preceeded by 11 bases with the sequence 5'-AAGAGG-3' which could form a ribosome binding site. Very few alternative translational signals exist in this region.

ORF	LENGTH AMINO ACIDS	MASS	CHARGE	pI	BASE COMPOSITION (A+T)
1	334	40.37 kDa	+13.71	9.01	74.65 %
2	166	18.99 kDa	+8.04	9.32	76.31 %
3	358	40.87 kDa	+10.14	8.83	75.76 %
4	45	5.30 kDa	+3.24	9.51	68.15 %
5	122	14.29 kDa	+3.23	8.87	73.50 %
6	47	5.09 kDa	+8.87	10.25	87.94 %
7	43	5.39 kDa	-0.15	6.26	75.97 %
8	392	46.2 kDa	+2.15	7.70	72.36 %

Table 6.8 Putative ORFs of pBP1

ORF'S 2, 3 and 4

Both ORF's 2 and 3 have putative methionine initiation codons preceded by sequences resembling ribosome binding sites (figure 6.3). The two genes are divergently arranged and their respective RBS are separated by 133 nucleotides. Recent evidence has indicated that transcripts are not thought to be functional if they start very close to the RBS (Vellanoweth, 1993); the full lengths of the promoters for these genes must be close if not overlapping. Analysis of the secondary structure of ORF's 2 and 3 using the Protean program of DNASTAR predicts that both putative proteins are largely hydrophilic in character. Their N-termini are strongly hydrophobic and on closer examination, these sequences appear to bear features typical of signal peptides (Izard & Kendall, 1994) suggesting an extra-cytoplasmic function for these peptides (figure 6.6).

ORF 3 contains 2 additional features. Close to its 5' end is a 15 bp motif [5'-GATTACCACAAACT-3'] which occurs as a tandem repeat, followed by a partial repeat and results in a amino acid motif DLPQT-DLPQT-DL. Whether this iteron is an important DNA or protein feature is unknown. However, the same amino acid motif could have been achieved with a range of codons. At the C-terminus is a hydrophobic region which could possibly function as a membrane anchor.

ORF 4 is located downstream of ORF 3, and its initiation codon overlaps with the stop codon of ORF 3 (figure 6.3). The secondary structure for this putative peptide predicts four β -pleated regions which are interspersed with turns; the hydropathicity profile suggests two hydrophilic regions alternating with three hydrophobic regions. Unlike ORF's 2 and 3, there is no evidence to suggest that ORF 4 possesses a signal sequence.

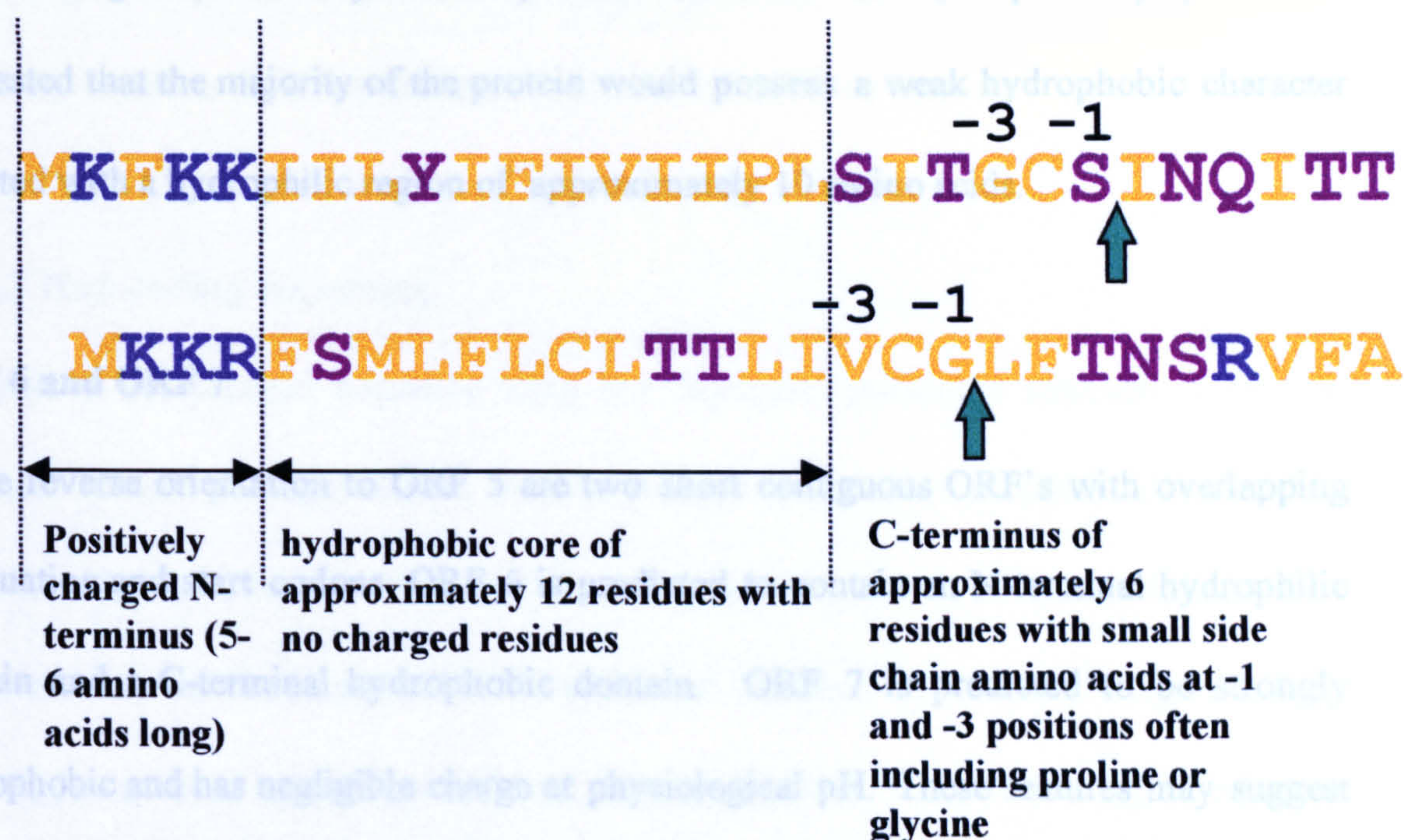


Figure 6.6 Putative signal peptides of ORF 2 and ORF 3 from pBP1

The N-terminus of ORFs 2 (top) and ORF 3 are annotated to highlight features which commonly occur in signal peptide sequences (Izard & Kendall, 1994). Positively charged amino acids are shaded in blue, negatively charged amino acids are shaded in red, neutral polar amino acids are shaded in purple whilst uncharged hydrophobic residues are shown yellow. Possible proteolytic cleavage sites are depicted by a green arrow and small amino acids (typically Alanine, Glycine, Serine or Threonine) at positions -1 and -3 to the cleavage site are also shown. These sequences contain N-terminal basic residues typical of signal peptides followed by a hydrophobic region which may extend slightly longer than 12 residues. The occurrence of some neutral polar residues within the proposed hydrophobic region is not inconsistent with this model as signal sequences are often 'sub-optimal' (Izard & Kendall, 1994).

ORF 5

This ORF has a methionine initiation codon preceded by a ribosome binding site, 5'-AGGAGG-3'. No similar proteins could be identified in a Swiss-Prot database search. Secondary structure analysis using the DNASTAR program 'Protean' predicted three α -helical regions, each separated by turns and coils. A hydropathicity prediction suggested that the majority of the protein would possess a weak hydrophobic character bisected with a hydrophilic region of approximately 10 amino acids.

ORF 6 and ORF 7

In the reverse orientation to ORF 5 are two short contiguous ORF's with overlapping termination and start codons. ORF 6 is predicted to contain an N-terminal hydrophilic domain and a C-terminal hydrophobic domain. ORF 7 is predicted to be strongly hydrophobic and has negligible charge at physiological pH. These features may suggest that these putative proteins are co-expressed and may be located in the membrane.

ORF 8

A comparison of the translated sequence for ORF 8 with the entire Swiss-Prot protein database revealed that this putative protein has a relatively large number of homologs, which are distributed principally among the various groups of RC plasmids (van der Lelie *et al.*, 1989). Whilst the majority of these have a methionine initiation codon, ORF 8 is predicted to start with a valine residue, specified by the alternative initiation codon GUG (figure 6.7). These proteins, which have been implicated in plasmid mobilisation (Priebe & Lacks, 1989), encode site-specific recombinases whose target sequence occurs within the plasmid, 5' to the recombinase gene (Pre). Overlapping this recombination

site, known as RS_A, are the promoter elements of the Pre protein. Comparison of the sequences in front of ORF 8 with characterised RS_A regions (van der Lelie *et al.*, 1989; Josson *et al.*, 1990) allowed the identification of a putative RS_A site and Pre promoter elements which are akin to the pUB110/pMV158 Pre group, but not with the pE194 group (figure 6.7). This sequence identity grouping was also reflected in the sequence similarity between the Pre proteins.

6.4.4.3 Non-coding sequences

Analysis of the plasmid sequence using the 'Squiggle' secondary structure prediction software (GCG package, Zuker, 1989) revealed a number of potential secondary structures. Inverted repeats typical of rho-independent transcriptional terminators were found between the converging ORF's 1 and 2 and between ORFs 4 and 5 (figure 6.3). A third region of dyad symmetry is found in between ORF 5 and 6 in a region which is presumed to encode the promoters for these genes. The A+T rich nature of these two structures, and their locations, are more likely to make them candidates for some aspect of transcriptional control or fortuitous features in the high A+T DNA content of this intergenic region (A+T=85.3 %). A further 2 inverted repeats are found 5' to the putative replication protein. One of these occurs 28 bp before a heptameric motif described as a double-stranded origin nick site (*dso*), characteristic of the pC194 group of plasmids (Seery *et al.*, 1993). Indeed, the occurrence of the combination of an inverted repeat, *dso* and pC194 type replication protein are indicative features that this extrachromosomal element is a plasmid which replicates via a RC mechanism. A second region of dyad symmetry is found between the *dso* and replication protein. Such features are found in

1	V	S	Y	L	V	C	T	M	K	K	M	K	K	D	N	L	F	G	L	Q	K	H	N	Q	R	E	S	E	N	H	ORF8	pBP1
1	M	S	Y	A	V	C	R	M	Q	K	V	K	S	A	G	L	K	G	M	Q	F	H	N	Q	R	E	R	K	S	R	Pre	pTB19
1	M	S	Y	L	V	A	N	M	Q	K	L	K	A	D	N	L	V	G	L	G	N	H	D	Q	R	R	T	Q	H	H	Mob	pLC88
1	M	S	Y	A	V	C	R	M	Q	K	V	K	S	A	G	L	K	G	M	Q	F	H	N	Q	R	E	R	K	S	R	Mob	pBC16

31	N	N	K	D	I	D	K	N	K	S	D	L	N	Y	D	L	V	N	E	N	K	I	N	L	S	E	K	V	F	D	ORF8	pBP1
31	T	N	D	D	I	D	H	E	R	T	R	E	N	Y	D	L	K	N	D	K	N	I	D	Y	N	E	R	V	K	E	Pre	pTB19
31	K	N	P	D	I	D	V	D	R	S	A	L	N	Y	D	L	V	A	G	R	T	G	H	F	K	T	D	I	E	A	Mob	pLC88
31	T	N	D	D	I	D	H	E	R	T	R	E	N	Y	D	L	K	N	D	K	N	I	D	Y	N	E	R	V	K	E	Mob	pBC16

61	L	I	E	K	L	R	E	S	E	K	A	I	R	K	D	A	V	L	L	D	E	F	I	I	S	S	D	R	D	F	ORF8	pBP1
61	I	I	E	S	Q	K	T	G	T	R	K	T	R	K	D	A	V	L	V	N	E	L	L	V	T	S	D	R	D	F	Pre	pTB19
61	Y	I	N	E	H	K	T	S	Q	R	A	V	R	K	D	A	V	L	V	N	E	W	I	I	S	S	D	S	H	F	Mob	pLC88
61	I	I	E	S	Q	K	T	G	T	R	K	T	R	K	D	A	V	L	V	N	E	L	L	V	T	S	D	R	D	F	Mob	pBC16

91	F	K	N	M	N	E	N	E	I	K	K	F	F	E	T	S	K	Q	F	F	A	E	R	Y	G	E	E	N	I	I	ORF8	pBP1
91	F	E	Q	L	D	P	G	E	Q	K	R	F	F	E	E	S	Y	K	L	F	S	E	R	Y	G	K	Q	N	I	A	Pre	pTB19
91	F	A	D	L	T	A	A	D	M	R	K	Y	F	E	T	A	K	D	Y	F	A	E	K	F	G	E	E	N	I	R	Mob	pLC88
91	F	E	Q	L	D	P	G	E	Q	K	R	F	F	E	E	S	Y	K	L	F	S	E	R	Y	G	K	Q	N	I	A	Mob	pBC16

121	S	A	N	V	H	L	D	E	T	T	P	H	M	H	L	D	V	C	P	L	R	N	G	-	R	L	-	S	K	N	ORF8	pBP1
121	Y	A	T	V	H	N	D	E	Q	T	P	H	M	H	L	G	V	V	P	M	R	D	G	-	K	L	Q	G	K	N	Pre	pTB19
121	Y	A	I	V	H	L	D	E	S	T	P	H	M	H	M	G	I	V	P	F	D	D	E	H	K	L	S	A	K	R	Mob	pLC88
121	Y	A	T	V	H	N	D	E	Q	T	P	H	M	H	L	G	V	V	P	M	R	D	G	-	K	L	Q	G	K	N	Mob	pBC16

149	I	F	T	R	N	-	F	L	R	-	N	E	E	L	P	K	-	L	K	N	N	G	F	F	L	E	R	G	V	K	ORF8	pBP1
150	V	F	N	R	Q	E	L	L	W	L	Q	D	K	F	P	E	H	M	K	K	Q	G	F	E	L	K	R	G	E	R	Pre	pTB19
151	V	F	N	R	T	A	L	R	D	V	Q	D	Q	L	P	A	Y	L	P	Q	H	G	L	N	I	Q	R	G	V	Q	Mob	pLC88
150	V	F	N	R	Q	E	L	L	W	L	Q	D	K	F	P	E	H	M	K	K	Q	G	F	E	L	K	R	G	E	R	Mob	pBC16

176 G F T - K H I D I Q I L K A R A E E K K I K E L - K E L K E ORF8 pBP1
180 G S D R K H I E T A K F K K Q T L E K E I D F L E K N L A V Pre pTB19
181 E S G R K S L A V P G Y K A - - - - - M R E S I K Q Mob pLC88
180 G S D R K H I E T A K F K K Q T L E K E I D L L E S E L K N Mob pBC16

204 K N K E I E L K I A R S K E K L E - - - - - E Y K D L D F Q ORF8 pBP1
210 K K D E W T A Y S D K V K S D L E V P A K R H M K S V E V P Pre pTB19
202 S Q Q K L A A V E N E T K - - - - - Q R Q A K L K T Y Q Mob pLC88
210 K K S E L A I L S E E V S G E F K I P V K R E K K S V E V P Mob pBC16

229 I K S K K T V L N N L N Q G N D C L D E D I K Y K L D G F D ORF8 pBP1
240 T G E K S M - - - - - F G L G K E I M K T E K K P T K N V V I Pre pTB19
225 A T K - - - - - F D V N S V K T K E S R F H K R Y V L V Mob pLC88
240 T G K R N L - - - - - L G I E Q K K T V M K S T G N V I L Mob pBC16

259 V K M D Y V E F R K F Q T Q I A S I K K V K K E - - N E D L ORF8 pBP1
266 S E R D Y K N L V T A A R D N D R L K Q H V R N L M S T D M Pre pTB19
248 D R F D F D K L K Q G A S - - - - - - - - - - - - - - Mob pLC88
264 K D E V F Q D L K Q K V K A G A V L Q T R V D Q L L N T D Y Mob pBC16

287 K N E N S I V H K K C S D L R K S Q E E V S S R F T N W I D ORF8 pBP1
296 A R E Y K K L S K E H G Q V K E K Y S G L V E R F N E N V N Pre pTB19
261 - L T Mob pLC88
294 A K E N Q S L K K E V K E L R S T N K S L S E - - - - - Mob pBC16

317 N T Y K H M S L D S R E Q F K R R G S E S L A N G S T G N - ORF8 pBP1
326 D - Y N E L - L E E N K S L K S K I S D - L K R D V S L I - Pre pTB19
263 D T Y F T E T L S Q R S D M D I Q K D Q L I K A E S K A M E Mob pLC88
317 - - - - - - - - - - E N G R L K S T V E H - L T N E I E S L - Mob pBC16

346	Y	D	E	D	S	T	L	D	K	S	V	M	D	S	-	K	A	F	N	E	-	-	D	V	Y	I	K	C	A	N	ORF8	pBP1
352	Y	E	S	T	K	E	F	L	K	E	R	T	D	G	L	K	A	F	K	N	V	F	K	G	F	V	D	K	V	K	Pre	pTB19
293	L	D	I	E	N	R	R	L	Q	K	L	V	G	T	L	Q	G	I	V	R	S	V	D	R	F	L	Q	R	K	L	Mob	pLC88
336	Y	V	L	T	K	D	F	L	K	A	R	T	N	D	L	E	S	F	K	E	L	F	G	V	F	V	G	K	V	K	Mob	pBC16

373	E	K	-	-	-	-	-	-	-	E	N	A	L	K	E	P	K	M	S	S	N	G	F	-	-	-	-	-	-	-	ORF8	pBP1
382	D	K	T	A	Q	-	-	F	Q	E	K	H	D	L	-	E	P	K	K	N	E	F	E	L	T	-	-	-	-	H	Pre	pTB19
323	G	V	G	L	P	S	E	W	L	E	R	A	G	L	K	E	P	S	K	K	A	P	-	-	-	-	-	-	Q	R	Mob	pLC88
366	E	K	A	P	R	G	L	F	V	R	Q	H	E	R	S	E	E	K	K	N	T	F	S	L	Q	D	V	L	Q	R	Mob	pBC16

389	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	E	I	D	W									ORF8	pBP1
405	N	R	E	V	K	K	E	R	S	-	-	-	-	-	R	D	Q	G	M	S	L										Pre	pTB19
347	P	Q	E	R	S	E	G	P	R	D	K	L	D	G	P	T	L	E	F	G	L										Mob	pLC88
396	D	R	E	L	R	E	Q	R	K	A	K	R	K	-	K	S	H	D	L	E	R										Mob	pBC16

Figure 6.7 Comparative alignment of ORF 8 from pBP1 with Pre/Mob proteins
 ORF 8 encoded on plasmid pBP1 from *C. botulinum* NCTC 2916 is compared with plasmid recombination (Pre) and mobilisation (Mob) proteins found on rolling circle plasmids from Gram-positive bacteria; pTB19 isolated from *Bacillus stearothermophilus* (van der Lelie *et al.*, 1989), pLC88 isolated from *Lactobacillus casei* (Bates & Gilbert, 1989) and pBC16 from *Bacillus cereus* (Tang & Wilson, unpublished data, EMBL-GenBank submission PPU32369)

other examples of this class of plasmid (Josson *et al.*, 1990; Devine *et al.*, 1989) and may also serve some function in plasmid maintenance/replication (figure 6.3).

The intergenic region between ORF 7 and the putative Pre gene contains a tract of DNA with a particularly high G+C base bias (47.3 % in 237 bp) in comparison to the remainder of the plasmid (25.66 % G+C). Inspection of this sequence revealed a proliferation of short inverted repeats which typically contained a high proportion of guanine and cytosine residues. This type of feature is found throughout the rolling circle group of plasmids and may function to initiate the synthesis of a second DNA strand (lagging strand) from a single stranded replicative intermediate (Dempsey *et al.*, 1995; Kramer *et al.*, 1997). Such motifs are commonly termed the single stranded origin (*sso*). A further inverted repeat is found spanning the 3' end of ORF 7; this could be a transcriptional termination signal associated with ORF 7 or participate in the proposed minus origin (figure 6.5).

6.5 Discussion

6.5.1 Transformation of group I *C. botulinum* type A strains

6.5.1.1 Evaluation of gene transfer in *C. botulinum* type A

Attempts to develop a gene transfer procedure for group I *C. botulinum* strains did not realise their potential as not all the criteria outlined in section 6.1.1 were achieved. Specifically, transformation frequency was not optimised to a point where gene knock out mutagenesis could be attempted with a suicide vector. The possible reasons for this are discussed.

6.5.1.2 Electroporation parameters

Attempts to optimise electrotransformation of Gram-positive bacteria have generally succeeded through the methodical alteration of parameters such as electrical field strength, resistance, plasmid concentration, cell density and time constants (Dunny *et al.*, 1991). Variations in these parameters are often accompanied by a concomitant fluctuation in transformants which may increase by the order of three to four logs. This increase was not observed with *C. botulinum* NCTC 2916, possibly because the appropriate parameters were not tried. Nevertheless, similar parameters have proved relatively successful for a number of other clostridia (Chen *et al.*, 1996; Phillips-Jones, 1995) including a type A *C. botulinum* strain (Zhou & Johnson, 1993). With these considerations in mind, it is unlikely that the electroporation parameters tested were entirely responsible for the poor transformation frequencies observed. Furthermore, they could not logically explain why one vector established in *C. botulinum* NCTC 2916 whilst a second vector did not. A similar argument for *C. botulinum* ATCC 3502 cannot be equally substantiated as less variables were tested.

6.5.1.3 Host encoded factors

Gram-positive organisms are in general, only poorly amenable to transformation, and one factor which may be significant is the production of extracellular non-specific nucleases. The inhibition of such nucleases has been implicated in the successful application of both protoplast transformation (Lin & Blasheck, 1984) and electrotransformation (Chen *et al.*, 1996) of some clostridial species. Indeed, both strains of type A *C. botulinum* used in this study as well as several non-proteolytic group II type B *C. botulinum* strains and *C. beijerinckii*, all exhibited DNase activity on DNase

agar (T. Davis unpublished observation). The results of Chen (1996) suggest that this type of factor could severely reduce transformation efficiency.

Restriction endonucleases are widespread amongst bacteria and have proven a considerable barrier to gene transfer, not least in a number of clostridia (Mermelstein & Papoutsakis, 1993; Azeddoug *et al.*, 1989; Azeddoug & Reysset, 1991; Chen *et al.*, 1996). Although restriction activity was not detected in group I *C. botulinum* in this study, the method may have been unsuitable, and for instance, disruption of the cells may have released endogenous proteases from these proteolytic clostridial species which could potentially inactivate a restriction endonuclease. Alternatively they could possess type I or type III restriction enzymes which have more specific substrate requirements than type II systems (Bickle *et al.*, 1987). Evidence of a restriction system has been demonstrated in a *C. sporogenes* strain (Dr M. Mauchline, pers. comm.) and characterised in a non-proteolytic type B *C. botulinum* strain (T. Davis, unpublished data). In these examples, the unprotected vector pGK12 could be introduced by electrotransformation, albeit at very low frequency (T. Davis, unpublished data). However, other plasmids such as pMTL540E could not. Two possible explanations could account for these findings. Firstly pGK12 has a low G+C DNA bias and hence contains few restriction sites in comparison with pMTL540E. Thus pGK12 may avoid restriction by virtue of its sequence. This may be the case in the non-proteolytic *C. botulinum* strain where pGK12 has a single target sequence [5'-CCGG-3'] whilst pMTL540E contained 14. Secondly, the mode of replication may be responsible. Plasmids such as pGK12 replicate via single stranded intermediates in both *E. coli* and Gram-positive hosts and it may be these single stranded intermediates which are inefficient substrates for type II restriction enzymes that are successfully introduced.

The fact that *C. botulinum* NCTC 2916 can only be transformed with pGK12 suggests that a restriction barrier may exist. This could also be true for *C. botulinum* ATCC 3502 if the restriction system recognised a low number of sites. As plasmid DNA can be purified from both strains, the methylation profile could be probed using a chemical modification method (Olek *et al.*, 1996).

An apparent increase in electrotransformation frequency by the addition of glycine to the growth media of *C. botulinum* ATCC 3502 suggests that the cell wall may also pose a barrier to the efficiency of the gene transfer process. Whether this is a physical barrier to the creation of 'pores', a disposition to rupture the whole cell, or an inability to arrest the chemical imbalances caused by pore formation causing lysis (Weaver & Chizmadzhev, 1996), is difficult to assess. The observation that very high kill rates are possible would suggest that the latter two hypotheses are more likely. Sub-inhibitory concentrations of glycine are thought to interfere with cell wall synthesis (Miller, 1994), so alterations to the electroporation protocol including the use of alternative cell wall modification agents and osmotic stabilizers, may be advantageous for these organisms.

Neurotoxigenic clostridia have been associated with an autolytic phenomenon and an autolytic enzyme has been characterised (Kawata & Takumi, 1971). The lytic enzyme was isolated from a type A strain during logarithmic growth which contained both endo- β -N-acetylglucosaminidase and N-acetylmuramyl-L-alanine amidase activities. In theory this enzyme would be present in the concentrated cell suspensions prepared for electrotransformation and would be liberated after electroporation when a large proportion of cells are damaged. It is conceivable that the autolysin would reduce the number of viable cells and, therefore, reduce the number of transformants.

6.5.1.4 Incompatibility

A possibility exists that plasmids indigenous to the strains used in this study already possess elements which are incompatible with the recombinant vectors. This situation typically arises when attempts are made to force plasmids with a similar replicon to coexist. As 'large scale' plasmid preparations were not conducted before gene transfer, the actual extra-chromosomal complement of these strains are unknown.

6.5.1.5 Conclusions

Gene transfer by electroporation has proved to be successful, albeit inefficient. Several possible barriers to this process have been postulated which could alone severely reduce transformation efficiency. In combination, these factors could pose a formidable barrier to the introduction of DNA to group I *C. botulinum* and must be individually addressed before an empirical assessment of physical parameters can be made. The introduction of a novel plasmid replicon into group I *C. botulinum* is an improvement on existing genetic systems which are currently limited to the use of 'single stranded' replicons and conjugative transposons (Zhou & Johnson, 1993; Lin, 1992; Lin & Johnson, 1991).

6.5.2 Theoretical characterisation of pBP1

6.5.2.1 Method of replication

The plasmid pBP1 contains several functional elements which enable its classification. The presence of a putative Rep protein and nick site which are homologous with members of the pC194 'rolling circle' (RC) plasmids (Seery *et al.*, 1993) strongly indicate that this is also a plasmid which replicates via this mechanism. Members of the RC plasmid family have been encountered as co-integrants with other plasmids

(Gennaro *et al.*, 1987). However, no evidence was found to support the existence of a second replicon in pBP1.

In addition to the requirement of initiating plasmid replication via the generation of a site specific 'nick' at the plus origin, ss intermediate forms of the plasmid must also be converted back to the ds form. DNA elements described as minus origins support this conversion but are less conserved than the plus origins. Different RC plasmids utilise independent types of minus origins which may be host specific (Seery *et al.*, 1993). Indeed, some plasmids, such as pMV158, may contain two or more separate types of minus origin (Khan, 1996). A region of complex dyad symmetry typical of minus origins has been detected in pBP1 which does not resemble *ssoA* or *ssoU* characterised in other plasmids (Kramer *et al.*, 1997). A recombinant version of pBP19 not containing this sequence was able to replicate autonomously in *C. beijerinckii* NCIMB 8052 (Dr M. Mauchline pers. comm.). The replication of plasmids without a functional minus origin has previously been described and suggests that host factors responsible for lagging strand synthesis from the minus origin may also be able to initiate from other sites within plasmids, albeit at a cost to plasmid stability and copy number (Dempsey *et al.*, 1995).

Plasmid replication control in the RC plasmids is thought to be proportional to the quantities of the initiator protein (Rep) synthesised. Thus, mechanisms which modulate Rep synthesis exert control on copy number. Many Rep proteins, including that of pC194, have antisense RNA transcripts which are able to negatively regulate Rep transcription by attenuating Rep mRNA synthesis, or regulate Rep translation by sequestering the ribosome binding site of the Rep transcript (Rasooly & Rasooly, 1997). On the basis of the similarity between ORF 1 of pBP1 and the Rep protein of pC194,

these two mechanisms would appear to be likely candidates in the regulation of pBP1 plasmid copy number. In addition, an unfavourable initiation codon (UUA) for the pBP1 Rep could also reduce the translational efficiency of this gene.

Members of other RC plasmid families may also have a further level of Rep protein regulation. Plasmids such as pMV158 and some other members of the pE194 family encode a trans-acting repressor termed 'Cop' which in the case of pMV158, negatively regulates Rep synthesis at the transcriptional level. At the current time, analogous Cop proteins have not been found in the pC194 family and may consequently be unlikely to exist in pBP1.

6.5.2.2 Pre protein function

A homologous group of proteins, termed Pre (plasmid recombination) or Mob proteins (mobilisation), commonly occur in RC plasmids. As their names suggest, they are thought to participate in plasmid recombination and mobilisation. Both activities are mediated via excision at a site specific sequence (RS_A), and can result in the formation of co-integrants with other plasmids with the appropriate RS_A site (Gennaro *et al.*, 1987), or in plasmid mobilisation in the presence of the appropriate transfer functions from conjugal plasmids (Priebe & Lacks, 1989). Unlike a number of unrelated site specific recombinases, Pre proteins do not appear to participate in the resolution of plasmid multimers (Gennaro *et al.*, 1987). The biological function of Pre in pBP1 and other plasmids is not completely clear, but its continued selection may result from its ability to form heterologous cointegrants which could enable concomitant transfer of the RC plasmid into other hosts, or to entrap other genes into the plasmid. Alternatively, it may function as an aid to assisted mobilisation.

6.5.2.3 Cassette structure

A number of medically important RC plasmids commonly found in organisms such as staphylococci also have closely related counterparts in other Gram-positive bacteria such as *B. subtilis* and *E. faecalis* (Novick, 1989). Genomic sequence analysis has revealed that certain common features, such as antibiotic resistance determinants, replicons and Pre recombinase genes, occur in a number of distinct plasmids. In essence, RC plasmids are composed of a mosaic structure of “cassettes” which can be divided into essential replicative structures (Rep, *dso*, *sso*, *cop*), mobilisation and recombinogenic structures (Pre, *mob*) and antibiotic resistance determinants. The abrupt nature with which these elements join suggests that precise ‘cassette exchange’ between plasmids has been responsible for plasmid evolution. In this light, examination of pBP1 also reveals a number of “cassettes”, including common features such as a replicon, Pre recombinase and minus origin. However, instead of encoding antibiotic resistance genes which would be of little use to the clostridium, this plasmid encodes six genes whose organisation resembles two regulons. Although their function remains cryptic, it is likely that ORFs 2 and 3 have an extracytoplasmic function whilst ORFs 4 and 5 may have a membrane location.

Comparison of pBP1 with two RC plasmids from other clostridia (Delver *et al.*, 1996; Brehm *et al.*, 1992) revealed that they contain replicons similar to the pC194 family and also possess ORFs encoding cryptic proteins. In addition, the secondary structure motifs presented as minus origins do not resemble the minus origins of staphylococcal and related plasmids. In conclusion, the structure of pBP1 would also appear to be a mosaic of structures with which we are already familiar, such as the type of replicon and recombinase gene. However, there are additional features which are not familiar. This is

perhaps not suprising when one considers, for instance, that the minus origin is thought to be host specific whilst the cryptic genes are likely to be pertinent to the environment in which the clostridium survive.

6.5.2.4 Natural distribution of pBP1

At least four investigations have documented the existence of plasmids indigenous to group I *C. botulinum* strains and their closely related non-toxigenic counterpart *C. sporogenes* (Scott & Duncan, 1978; Strom *et al.*, 1984; Weickert *et al.*, 1986; Ferriera *et al.*, 1987). The most comprehensive study of type A strains was conducted by Weickert and co-workers who reported the presence of 5 individual plasmids which were apparently distributed to different degrees in 12 distinct type A isolates, and in some cases, in *C. sporogenes*. These results are somewhat difficult to interpret as they are of ccc forms which generally migrate faster than linear or oc forms.

An earlier study examined type A organisms and also found that distinct plasmid species appeared to occur in a number of isolates (Strom *et al.*, 1984). The size of the ccc forms were corroborated with restriction digests using λ /Hind III fragments as molecular weight markers. Plasmids of approximately 5.0 kb, 9.7 kb, 18.4 kb and 38 kb were encountered, but none of the sizes reported appeared to correlate closely with the size of pBP1. The size of these plasmids roughly equated with the findings of Weickert (1986), Ferreira (1987) and with one of the plasmids described by Scott and Duncan (1978), which may indicate that certain plasmids are widely distributed in this organism which may be clonal. However, when all the current toxigenic members of group I *C. botulinum* are considered, a range of plasmid species are present, suggesting that the extrachromosomal complements of this group are not clonal. Furthermore, 16S rRNA

gene sequencing (Collins *et al.*, 1994) has demonstrated that the group I *C. botulinum* and *C. sporogenes* are not identical, whilst genomic fingerprinting has shown that *C. botulinum* type A strains are also not clonal (Lin & Johnson, 1995).

In conclusion, there is no firm evidence to suggest pBP1 is widely distributed among the type A *C. botulinum* strains or their close relatives.

6.5.2.5 Potential uses

To date no other plasmids have been cloned from group I *Clostridium botulinum* strains, a factor which has previously made heterologous plasmids the vector of choice for studies in these organisms (Zhou & Johnson, 1993). Sequence analysis of pBP1 will now allow a rational approach to the construction of a shuttle vector using the pBP1 replicon. However, the similarity of the putative pBP1 replicon with the pC194 family of plasmids strongly implies that this plasmid replicates via a RC mechanism, which are considered intrinsically unstable (Kiewiet *et al.*, 1993; Janniere *et al.*, 1990). Nonetheless, an indigenous plasmid may prove useful for cloning relatively small DNA inserts, particularly if the plasmid design sympathetically considers the natural order of pBP1.

Relatively few genes have been cloned and studied from *C. botulinum* strains. Thus, pBP1 may also be an interesting source of transcriptional and translational signals. The prediction of two signal peptide sequences for ORFs 2 and 3 could be useful for the design of recombinant proteins that require export across the cytoplasmic membrane, such as β -lactamases, which would increase the repertoire of resistance determinants for these organisms.

CHAPTER 7

GENERAL DISCUSSION

7.1 Introduction

Genetic analysis of botulinal neurotoxin operons has received a substantial amount of attention from the scientific community, partly because they continue to pose a threat to human health, particularly with the constant consumer pressure to reduce the processing/preservatives of the foods we eat. The neurotoxins (in the form of a neurotoxin complex) have also demonstrated their value as a tool in health care for the treatment of movement disorders. Both of these areas have considerable economic importance. The specific properties of the neurotoxin complex, particularly in host cell targeting are likely to continue this trend.

The data generated in this study should contribute to the understanding of neurotoxin complex production and pave the way for a comprehensive study of factors which regulate neurotoxin production. Knowledge in both of these areas will be directly useful in the attainment of high neurotoxin titres for pharmaceutical applications and combat the production of neurotoxins in foods.

7.2 The type B neurotoxin gene in *C. botulinum* NCTC 2916 type A

Genetic characterisation of the type B neurotoxin gene in this strain revealed a number of anomalies which are capable of completely disrupting productive translation. However, it appeared that the pseudogene was still actively transcribed. Although these studies reveal how this has occurred, they do not answer why.

Clues which may resolve this question are found in the observation that the incidence of silent type B neurotoxin genes in type A *C. botulinum* strains is approximately 50 %, which is much higher than the incidence of dual neurotoxin producing strains (Franciosa *et al.*, 1994). This could be interpreted as a selective advantage.

Secondly the conclusion that the nontoxic/haemagglutinin genes found adjacent to the silent type B pseudogene are unique in this strain, and that no haemagglutinin genes are to be found 5' to the type A neurotoxin gene, suggests an active role for the type B nontoxic/haemagglutinin proteins.

Finally a survey of recent literature on the composition of the regulons accompanying neurotoxin genes suggests a bifurcation of genes and respective proteins which constitute neurotoxin complexes (Kubota *et al.*, 1998; Hutson *et al.*, 1996).

The basis of this classification is whether a haemagglutinin operon is present. 'Infant' type A₂, type E and F strains do not possess haemagglutinin operons but have a gene termed p47. In contrast, classical type A₁ strains and type B, C, D and G strains do (Kubota *et al.*, 1998; Bhandari *et al.*, 1997). The comparisons are not completely direct as there is sequence divergence between the genes, rearrangements (i.e, position and orientation of ORF 3 homologues) and deletions/insertions (i.e, type G has no HA 33 gene). In summary strains such as NCTC 2916 appear to possess a type B pseudogene in combination with a NTNH gene and haemagglutinin operon. In addition they accommodate a type A₁ neurotoxin gene with a NTNH gene and a 'p47' gene which closely resembles the situation in 'infant' type A₂ strains, although these latter strains have a type A₂ neurotoxin gene and no silent type B/haemagglutinin regulon.

The reason for this could be that the strains with haemagglutinins are more proficient pathogens as the haemagglutinins facilitate the intestinal absorption of toxin (for example, L and LL forms produced by A₁, B, C and D strains are considerably more toxic than M complexes produced by A₂, E and F strains (Fujinaga *et al.*, 1997).

Thus, type A strains with the type B pseudogene and haemagglutinins located only at the type B regulon will have the advantage of having a type A progenitor complex containing haemagglutinins in addition to what ever advantage is conferred by p47.

7.3 Recombination in *Clostridium botulinum*

The discovery that very similar neurotoxin genes are linked to disparate genes and operons that encode other components of the neurotoxin complex (i.e., NTN_H and haemagglutinin genes), and as yet uncharacterised genes of cryptic function (i.e. p47 genes), presents a very strong case for recombination. Indeed, the discovery of certain hybrid NTN_H genes and hybrid neurotoxin genes indicates that interspecies gene transfer events must be occurring between unrelated species (Kubota *et al.*, 1996; East *et al.*, 1996; Moriishi *et al.*, 1996). Presumably there are more combinations of neurotoxin regulons to be discovered which may possibly be found in a wider spectrum of species than are already recognised as being neurotoxigenic. It is interesting to note that neurotoxigenic strains of *C. baratii* are quite closely related to *C. perfringens*! (Collins *et al.*, 1994). Strains represented by *C. botulinum* NCTC 2916 would seem to represent a stable transition between the two common examples of type A strains. One could speculate that the p47-containing operon found adjacent to the type A neurotoxin gene was involved in a recombinational event with the silent type B locus resulting in the subsequent loss of the silent type B pseudogene locus and exchange of the haemagglutinin components to the type A gene. Such an event would result in a neurotoxin complex regulon which is virtually indistinguishable from the classical type A₁ strains.

7.4 Regulation of neurotoxin complex production in *Clostridium botulinum*

The association of the silent type B neurotoxin gene with a contiguous NTN_H gene and opposing haemagglutinin genes is an organisation which appears to be reflected in all known haemagglutinin positive strains of *C. botulinum* (summarised in Kubota *et al.*, 1998). The closely situated genes with a common function bear the hall marks of a co-

ordinately expressed regulon. The combination of both conserved promoter elements and putative transcription factor which occurs in various strains of *C. botulinum* (Henderson *et al.*, 1995), *C. tetani* and other *Clostridium*, support this conclusion. However, little light has been shed on what type of transcriptional factor ORF 3 is (i.e, alternative sigma factor/phosphorylated response regulator/transcriptional activator). Moreover, even less is known about what factors might interact with ORF 3 at the transcriptional or post translational level, and how these may correlate with the availability of exogenous carbon and nitrogen sources (Patterson-Curtis & Johnson, 1989; Leyer & Johnson, 1990).

The presence of the p47 protein, in what appears to be an operon contiguous with the NTNH gene and neurotoxin gene in type A₁/B^{*}, A₂, E and F strains, is a mystery as it does not appear to be associated with the neurotoxin complexes (Sugiyama, 1980). These operons do however possess a homologue of ORF 3 and a similarly conserved promoter element, suggesting that they may also be co-ordinately regulated.

7.5 Development of a genetic system ~ critical evaluation

During the course of this study a series of novel plasmid vectors based on the replicon of pCB102 were constructed and subsequently shown to function in several diverse clostridial species. These included a representative of group I *C. botulinum*, *C. perfringens* and several species of solventogenic clostridia. These findings indicate that vectors based on the replicon of pCB102 may be generally useful for use in 'cluster one' of the clostridia (Collins *et al.*, 1994). Preliminary data would indicate that the pCB102-derived vectors tested were both structurally and segregationally stable. Furthermore, sequence analysis of the minimal replicon implied that pCB102 did not belong to the commonly encountered classes of single-stranded replicating plasmids, but had some characteristics which were generally seen in other types of plasmids, e.g, the potential for

a 'kissing complex' which is seen in ColE1 or pAM β 1-like plasmid replicons. However, a study of the nucleic acid and protein replicative apparatus would be necessary to support this hypothesis. The availability of a structurally stable plasmid vector for use in *C. botulinum* type A is an improvement on the existing class of single-stranded replicating plasmid vector currently in use (Zhou & Johnson, 1993; Janniere *et al.*, 1990). This should facilitate the implementation of *in vivo* reporter expression and complementation studies. Nevertheless, unless a conditionally selective replicon can be constructed (for instance a temperature sensitive replicon of a pCB102 derived vector) a significant improvement in gene transfer frequency is required to generate chromosomal 'knockout' mutants via a suicide vector approach. Improvements in gene transfer frequency by electroporation are likely to involve a more thorough assessment of factors, including the contribution of the cell wall, buffer composition, non-specific nucleases and restriction endonucleases. The presence of a restriction system in NCTC 2916 seems likely due to the successful transformation of one plasmid and not another. However, these could be more complex than type II restriction systems and, therefore, more difficult to detect. Efficient gene transfer may be impeded by poor recognition of the heterologous plasmid replicative signals, in this case plasmid pBP1, isolated and characterised in this study may prove useful. Nucleotide sequence analysis of pBP1 would strongly suggest that it replicates via a 'single stranded' mechanism. As previously mentioned, this type of plasmid is not ideal for use as a recombinant vector. In hindsight, characterisation of a larger indigenous plasmid (>10 kb) may have yielded a more generally useful plasmid replicon.

REFERENCES

- Adachi, T., Yamagata, H., Tsukagoshi, N. and Udaka, S.** 1989. Multiple and tandemly promoters of the cell wall protein gene operon in *Bacillus brevis* 47. J. Bacteriol. **171**:1010-1016.
- Agaisse, A. and Lereclus, D.** 1996. STAB-SD: a Shine-Dalgarno sequence in the 5' untranslated region is a determinant of mRNA stability. Mol. Microbiol. **20**(3):633-643.
- Ahnert-Hilger, G. and Bigalke, H.** 1995. Molecular aspects of tetanus and botulinum neurotoxin poisoning. Progress in Neurobiol. **46**:83-96.
- Aktories, K., Selzer, J., Hofmann, F. and Just, I.** 1997. Molecular mechanism of action of *Clostridium difficile* toxins A and B. In The Clostridia: Molecular Biology and Pathogenesis. P.393-407 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Allen, S.P. and Blaschek, H.P.** 1988. Electroporation induced transformation of intact cells of *Clostridium perfringens*. Appl. Environ. Microbiol. **54**:2322-2324.
- Alt-Morbe, J., Stryker, J.L., Fuqua, C., Li, P-L., Farrand, S.K. and Winans, S.C.** 1996. The conjugal transfer system of *Agrobacterium tumefaciens* octopine-type Ti plasmids is closely related to the transfer system of an IncP plasmid and distantly related to Ti plasmid *vir* genes. J.Bacteriol. **178**(14):4248-4257.
- Anderson, A.S., Pettersson, A.F. and Kjeldsen, T.B.** A fast and simple technique for sequencing plasmid DNA with sequenase® using heat denaturation. Biotechniques. **13**(5):678-679.

- Arnon, S.S.** 1997. Human tetanus and human botulism. In *The Clostridia: Molecular Biology and Pathogenesis*. P.95-115 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Atherton, J.C., Cao, P., Peek Jr., R.M., Tummuru, M.K.R., Blaser, M.J. and Cover, T.L.** 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. *J. Biol. Chem.* 270(30):17771-17777.
- Aureli, P., Fenicia, B., Pasolini, M., Gianfranceschi, L.M., McCroskey, L.M. and Hatheway, C.L.** (1986) Two cases of type E infant botulism caused by neurotoxigenic *Clostridium butyricum* in Italy. *J. Infect. Dis.* 154:201-206.
- Azeddoug, H. and Reyset, G.** 1989. Characterisation of a methyl-specific restriction system in *Clostridium acetobutylicum* strain N1-4081. *FEMS Microbiol. Letts.* 65:323-326.
- Azeddoug, H. and Reyset, G.** 1991. Recognition sequence of a new methyl-specific restriction system from *Clostridium acetobutylicum* strain ABKn8. *FEMS Microbiol. Letts.* 78:153-156.
- Azeddoug, H., Hubert, J. and Reyset, G.** 1992. Stable inheritance of shuttle vectors based on plasmid pIM13 in a mutant strain of *Clostridium acetobutylicum*. *J. Gen. Microbiol.* 138:1371-1378.
- Barne, K.A., Bown, J.A., Busby, J.W. and Minchin, S.D.** 1997. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended -10' motif at promoters. *EMBO J.* 16(13):4034-4040.
- Bechhofer, D.H. and Zen, K.H.** 1989. Mechanism of erythromycin-induced *ermC* mRNA stability in *Bacillus subtilis*. *J. Bacteriol.* 171(11):5803-5811.

- Belasco, J.G. and Higgins, C.F.** 1988. Mechanisms of mRNA decay in bacteria: a perspective. *Gene*. 72:15-23.
- Bhandari, M., Campbell, K.D., Collins, M.D. and East, A.K.** 1997. Molecular characterisation of the clusters of genes encoding the botulinum neurotoxin complex in *Clostridium botulinum* (*Clostridium argentinense*) type G and nonproteolytic *Clostridium botulinum* type B. *Curr. Microbiol.* 35(4):207-214.
- Bhattacharyya, S.D. and Sugiyama, H.** 1989. Inactivation of botulinum and tetanus toxins by chelators. *Infect. Immun.* 57(10):3053-3057.
- Bhavsar, D., Zheng, Hd. And Drysdale, J.** 1994. Chimerism in PCR products from a multigene family. *Biochem. Biophys. Res. Comm.* 205(1):944-947.
- Bickle, T. A.** 1987. DNA restriction and modification systems. In *Escherichia coli* and *Salmonella typhimurium*. P.692-695 Eds. Neidhardt, F. C. ASM Washington DC.
- Binz, T., Kurazono, H., Popoff, M.R., Eklund, M.W., Sakaguchi, G., Kozaki, S., Kreigstein, K., Henschen, A., Gill, D.M. and Niemann, H.** 1990. Nucleotide sequence of the gene encoding *Clostridium botulinum* neurotoxin type D. *Nucl. Acids Res.* 18(18):5556.
- Binz, T., Kurazono, H., Wille, M., Frevert, J., Wernars, K. and Niemann, H.** 1990. The complete sequence of botulinum neurotoxin type A and comparsion with other Clostridial neurotoxins. *J. Biol. Cehm.* 265(16):9153-9158.
- Birnboim, H.C. and Doly, J.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* 7(6):1513-1523.
- Bittner, M.A., DasGupta, B.R. and Holz, R.W.** 1989. Isolated light chains of botulinum neurotoxins inhibit exocytosis. *J. Biol. Chem.* 264(18):10354-10360.

- Bjornsson, A. and Isaksson, L.A.** 1993. UGA codon context which spans three codons: reversal by ms²i⁶A37 in tRNA, mutation in *rpsD*(S4) or streptomycin. *J. Mol. Biol.* **232**:1017-1029.
- Black, J.D. and Dolly, J.O.** 1986a. Interaction of ¹²⁵I-labelled botulinum neurotoxins with nerve terminals. I. Ultrastructural autoradiographic localisation and quantitation of distinct membrane acceptors for types A and B on motor nerves. *J. Cell. Biol.* **103**:521-534.
- Black, J.D. and Dolly, J.O.** 1986b. Interaction of ¹²⁵I-labelled botulinum neurotoxins with nerve terminals. II. Autoradiographic evidence for its uptake into motor nerves by acceptor-mediated endocytosis. *J. Cell Biol.* **103**:535-544.
- Blasi, J., Chapman, E.R., Link, E., Binz, T., Yamasaki, S., De Camilli, P., Sudhof, T. C., Niemann, H. and Reinhard, J.** 1993. Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25. *Nature.* **365**:160-163
- Blasi, J., Chapman, E.R., Yamasaki, S., Binz, T., Neimann, H. and Jahn, R.** 1993. Botulinum neurotoxin C1 blocks neurotransmitter release by means of cleaving HPC-1/syntaxin. *EMBO J.* **12**:4821-4828
- Blaustein, R.O., Germann, W.J., Finkelstein, A and DasGupta, B.R.** 1987. The N-terminal of the heavy chain of botulinum type A neurotoxin forms channels in planar phospholipid bilayers. *FEBS lett.* **226**(1):115-120.
- Boe, L. and Rasmussen, K.V.** 1996. Suggestions as to quantitative measurements of plasmid loss. *Plasmid.* **36**:153-159.
- Bonventre, P.F. and Kempe, L.L.** 1960. Physiology of toxin production by *Clostridium botulinum* types A and B. *J. Bacteriol.* **79**:18-23.

- Bonventre, P.F.** 1979. Absorption of botulinal toxin from the gastrointestinal tract. *Rev. Infect. Dis.* 1(4):663-667.
- Bonventre, P.F. and Kempe, L.L.** 1959a. Physiology of toxin production by *Clostridium botulinum* types A and B. II. Effect of carbohydrate source on growth, autolysis and toxin production. *J. Bacteriol.* 76:372-374.
- Bonventre, P.F. and Kempe, L.L.** 1959b. Physiology of toxin production by *Clostridium botulinum* types A and B. II. Effect of pH and temperature during incubation on growth, autolysis and toxin production. *J. Bacteriol.* 76:374-377.
- Boquet, P. and Duflot, E.** 1982. Tetanus toxin fragment forms channels in lipid vesicles at low pH. *Proc. Natl. Acad. Sci. USA.* 79:7614-7618.
- Boyer, H.W. and Roulland-Dussoix, D.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. *J. Mol. Biol.* 49:459-463
- Brantl, S and Wagner, E.G.H.** 1996. An unusually long lives antisense RNA in plasmid copy number control: *in vivo* RNAs encoded by the Streptococcal plasmid pIP501. *J. Mol. Biol.* 255:275-288.
- Brantl, S.** 1994. The *copR* gene product of plasmid pIP501 acts as a transcriptional repressor at the essential *repR* promoter. *Mol. Microbiol.* 14(3):473-483.
- Brantl, S., Birch-Hirschfeld, E. and Behnke, D.** 1993. RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. *J. Bacteriol.* 175(13):4052-4061.
- Bravo, A.** 1997. Phylogenetic relationships of *Bacillus thuringiensis* δ -endotoxin family proteins and their functional domains. *J. Bacteriol.* 179(9):2793-2801.

- Brehm, J.K., Oultram, J.D., Thompson, D.E., Swinfield, T.J., Peck, H., Young, M. and Minton, N.P.** 1988. Construction of plasmid vector systems for *Clostridium acetobutylicum*. In Genetics and Biotechnology of Bacilli. Volume 2. p.409-414. Eds. Ganesan, A.T. and Hoch, J.A. Academic Press Inc.
- Brehm, J.K., Pennock, A., Bullman, H.M.S., Young, M., Oultram, J.D. and Minton, N.P.** 1992. Physical characterisation of the replication origin of the cryptic plasmid pCB101 isolated from *Clostridium butyricum* NCIB 7423. Plasmid. 28:1-13.
- Brown, C.M., Stockwell, P.A., Trotman, C.N.A. and Tate, W.P.** 1990. The signal for the termination of protein synthesis in prokaryotes. Nucl. Acids. Res. 18:2079-2086.
- Bruand, C., Ehrlich, S.D. and Janniere, L.** 1991. Unidirectional theta replication of the structurally stable *Enterococcus faecalis* plasmid pAM β 1. EMBO J. 10:2171-2177.
- Bruand, C., Le Chatelier, E., Ehrlich, D.S. and Janniere, L.** 1993. A fourth class of theta-replicating plasmids: The pAM β 1 family from Gram-positive bacteria. Proc. Natl. Acad. Sci. USA. 90:11668-11672.
- Burns, D.M., Burger, M.J. and Beacham, I.R.** 1995. Silent genes in bacteria: the previously designated 'cryptic' *ilvHI* locus of '*Salmonella typhimurium* LT2' is active in natural isolates. FEMS Microbiol. Letts. 131:167-172.
- Byeon, W-H. and Weisblum, B.** 1990. Replication genes of plasmid pE194-*cop* and *repF*: Transcripts and encoded proteins. J.Bacteriol. 172(10):5892-5900.
- Caldwell, A.L. and Gulig, P.A.** 1991. The *Salmonella typhimurium* virulence plasmid encodes a positive regulator of a plasmid-encoded virulence gene. J. Bacteriol. 173(22): 7176-7185.

- Campbell, K.D., Collins, M.D. and East, A.K.** 1993. Gene probes for identification of the botulinal neurotoxin gene and specific identification of neurotoxin types B, E and F. *J. Clin. Microbiol.* **31**(9):2255-2262.
- Cato, E.P., George, W.L. and Finegold, S.M.** 1986. Genus *Clostridium*, Bergey's manual of systematic bacteriology, vol.2. P.1141-1200. P.H.A. Sneath, N.S. Mair, M.E. Sharpe, and J.G. Holt (ed.), Williams and Wilkins, Baltimore.
- Cha, J., Bishai, W. and Chandrasegaran, S.** 1993. New vectors for direct cloning of PCR products. *Gene.* **136**:369-370.
- Chambers, S. P., Prior, S. E., Barstow, D. A. and Minton, N. P.** 1988. pMTL *nic*⁻ cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene.* **68**:139-149
- Chapman, E.R., Blasi, J., An, S., Brose, N., Johnston, P.A., Sudhof, T.C. and Jahn, R.** 1996. Fatty acylation of synaptotagmin in PC12 cells and synaptosomes. *Biochem. Biophys. Res. Comm.* **225**:326-332.
- Chen, C-K., Boucle, C. M. and Blasheck, H. P.** 1996. Factors involved in the transformation of previously non-transformable *Clostridium perfringens* type B. *FEMS Microbiol. Lett.* **140**:185-191.
- Chen, E.Y. and Seeburg, P.H.** 1985. Supercoil sequencing: A fast and simple method for sequencing plasmid DNA. *DNA.* **4**:165-170
- Chen, L-M., Hobbie, S. and Galan, J.E.** 1996. Requirement of CDC42 for *Salmonella*-induced cytoskeletal and nuclear responses. *Science.* **274**:2115-2118

- Clewell, D.B. and Helinski, D.R. 1969.** Supercoiled circular DNA-protein complex in *Escherichia coli*: Purification and induced conversion to an open circular DNA form. *Biochem.* **62**:1159-1166.
- Climo, M.W., Sharma, V.K. and Archer, G.L. 1996.** Identification and characterisation of the origin of conjugative transfer (*oriT*) and a gene (*nes*) encoding a single-stranded endonuclease on the staphylococcal plasmid pGO1. *J.Bacteriol.* **178**(16):4975-4983.
- Cohen, S.N., Chang, A.C.Y. and Hsu, L. 1972.** Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci. USA.* **69**:2110-
- Cole, S. T. and Canard, B. 1997.** Structure, organisation and evolution of the genome of *Clostridium perfringens*. In *The Clostridia: Molecular Biology and Pathogenesis*. P.49-63 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Coleman, J.E. 1992.** Zinc proteins: enzymes, storage proteins, transcription factors and replication proteins. *Ann. Rev. Biochem.* **61**:897-946
- Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. and Farrow, J.A.E. 1994.** The phylogeny of the genus *Clostridium*; Proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* **44**(4):812-826.
- Currie, M.G., Fok, K.F., Kato, J., Moore, R.J., Hamra, F.K., Duffin, K.L. and Smith, C.E. 1992.** Guanylin: An endogenous activator of intestinal guanylate cyclase. *Proc. Natl. Acad. Sci.* **89**:947-951

- Czeczulin, J.R., Hanna, P.C. and McClane, B.A.** 1993. Cloning, nucleotide sequencing and expression of the *Clostridium perfringens* enterotoxin gene in *Escherichia coli*. *Infect. Immun.* **61**:3429-3439.
- Daniels-Holgate, P.U. and Dolly, J.O.** 1996. Productive and non-productive binding of botulinum neurotoxin A to motor nerve endings are distinguished by its heavy chain. *J. Neurosci. Res.* **44**:263-271.
- DasGupta, B.R. and Boroff, D.A.** 1968. Separation of toxin and hemagglutinin from crystalline toxin of *Clostridium botulinum* type A by anion exchange chromatography and determination of their dimensions by gel filtration. *J. Biol. Chem.* **243**:1065-1072.
- DasGupta, B.R. and Sugiyama, H.** 1972. A common subunit structure in *Clostridium botulinum* type A, B and E toxins. *Biochem. Biophys. Res. Comm.* **48**(1):108-112.
- DasGupta, B.R., Boroff, D.A. and Rothstein, E.** 1966. Chromatographic fractionation of the crystalline toxin of *Clostridium botulinum* type A. *Biochem. Biophys. Res. Comm.* **22**:750-756
- De Chateau, M. and Bjorck, L.** 1994. Protein PAB, a mosaic albumin-binding bacterial protein representing the first contemporary example of module shuffling. *J. Biol. Chem.* **269**(16):12147-12151.
- del Solar, G., Kramer, G., Ballester, S. and Espinosa, M.** 1993. Replication of the promiscuous plasmid pLS1: a region encompassing the minus origin of replication is associated with stable plasmid inheritance. *Mol. Gen. Genet.* **241**:97-105.
- Delver, E.P., Belogurova, N.G., Tupikova, E.E., Varfolomeyev, S.D. and Belogurov, A.A.** 1996. Characterisation, sequence and mode of replication of plasmid pNB2 from the thermophilic bacterium *Clostridium thermosaccharolyticum*. *Mol. Gen. Genet.* **253**:166-172.

- Dempsey, L.A., Zhao, A.C. and Khan, S.A.** 1995. Localisation of the start sites of lagging-strand replication of rolling-circle plasmids from Gram-positive bacteria. *Mol. Microbiol.* **15**(40):679-687.
- Dertzbaugh, M.T. and West, M.W.** 1996. Mapping of protective and cross-reactive domains of the type A neurotoxin of *Clostridium botulinum*. *Vaccine.* **14**(16):1538-1544.
- Devine, K.M., Hogan, S.T., Higgins, D.G. and McConnell, D.J.** 1989. Replication and segregational stability of *Bacillus* plasmid pBAA1. *J. Bacteriol.* **171**(2):1166-1172.
- DiMari, J.F. and Bechhofer, D.H.** 1993. Initiation of mRNA decay in *Bacillus subtilis*. *Mol. Microbiol.* **7**(5):705-717.
- Dodd, I.B. and Egan, J.B.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucl. Acids. Res.* **18**(17):5019-5026.
- Doellgast, G.J., Beard, G.A., Bottoms, J.D., Cheng, T., Roh, B.H., Roman, M.G., Hall, P.A. and Triscott, M.X.** 1994. Enzyme-linked immunosorbent assay and enzyme-linked coagulation assay for detection of *Clostridium botulinum* neurotoxins A, B and E and solution-phase complexes with dual-label antibodies. *J. Clin. Microbiol.* **32**(1):105-111.
- Dolly, J.O., Ashton, A.C., McInnes, C., Wadsworth, J.D.F., Poulain, B., Tauc, L., Shone, C.C. and Melling, J.** 1990. Clues to the multi-phasic inhibitory action of botulinum neurotoxins on release of transmitters. *J. Physiol. (Paris).* **84**:237-246.
- Dombroski, A.J., Johnson, B.D., Lonetto, M. and Gross, C.A.** 1996. The sigma subunit of *Escherichia coli* RNA polymerase senses promoter spacing. *Proc. Natl. Acad. Sci. USA.* **93**:8858-8862.

- Dombroski, A.J., Walter, W.A., Record Jr., M.T., Siegele, D.A. and Gross, C.A.** 1992. Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell*. 70:501-512.
- Donovan, J.J.** 1986. Ion-conducting channels produced by botulinum toxin in planar lipid membranes. *Biochem.* 25:2872-2876.
- Donovan, W.P., Rupa, M.J., Slaney, A.C., Malvar, T., Gawron-Burke, M.C. and Johnson, T.B.** 1992. Characterisation of two genes encoding *Bacillus thuringiensis* insecticidal crystal proteins toxic to *Coleoptera* species. *Appl. Environ. Microbiol.* 58:3921-3927.
- Drams, S., Kocks, C., Forestier, C. and Cossart, P.** 1993. Internalin-mediated invasion of epithelial cells by *Listeria monocytogenes* is regulated by the bacterial growth of state, temperature and the pleiotropic activator *prfA*. *Mol. Microbiol.* 9:931-941.
- Drolet, M., Zanga, P. and Lau, P.C.K.** 1990. The mobilisation and origin of transfer regions of a *Thiobacillus ferrooxidans* plasmid: relatedness to plasmids RSF1010 and pSC101. *Mol. Microbiol.* 4(8):1381-1391
- Dunny, G.M., Lee, L.N. and LeBlanc, D.J.** 1991. Improved electroporation and cloning vector system for Gram-positive bacteria. *Appl. Environ. Microbiol.* 57(4):1194-1201.
- Dybvig, K. and Woodward, A.** 1992. Cloning and DNA sequence of a mycoplasmal *recA* gene. *J. Bacteriol.* 174(3):778-784.
- Dziejman, M. and Mekalanos, J.J.** 1995. Two component signal transduction and its role in the expression of bacterial virulence factors. In *Two-Component Signal Transduction*. P. 305-317. Ed. Hoch, J.A and Silhavy, T.J. American Society for Microbiology, Washington DC.

- East, A.K. and Collins, M.D. 1994.** Conserved structure of genes encoding components of botulinum neurotoxin complex M and the sequence of the gene coding for the nontoxic component in nonproteolytic *Clostridium botulinum* type F. *Curr. Microbiol.* **29**:69-77.
- East, A.K., Stacey, J.M. and Collins, M.D. 1994.** Cloning and sequencing of a hemagglutinin component of the botulinum neurotoxin complex encoded by *Clostridium botulinum* types A and B. *System. Appl. Microbiol.* **17**:306-312.
- Eguchi, Y., Itoh, T. and Tomizawa, J-I. 1991.** Antisense RNA. *Ann. Rev. Biochem.* **60**:631-652.
- Ehretsmann, C.P., Carpousis, A.J. and Krish, H.M. 1992.** mRNA degradation in procaryotes. *FASEB J.* **6**:3186-3192.
- Eisel, U., Jarausch, W., Goretzki, K., Henschyen, A., Engels, J., Weller, U., Hudel, M., Habermann, E. and Niemann, H. 1986.** Tetanus toxin: primary structure, expression in the *E. coli* and homology with botulinum toxins. *EMBO J.* **5**(10):2495-2502.
- Eklund, M.W. and Poysky, F.T. 1974.** Interconversion of type C and D strains of *Clostridium botulinum* by specific bacteriophages. *Appl. Microbiol.* **27**(1):251-258.
- Eklund, M.W., Poysky, F.T. and Reed, S.M. 1972.** Bacteriophage and the toxigenicity of *Clostridium botulinum* type D. *Nature New Biology.* **235**:16-17.
- Eklund, M.W., Poysky, F.T., Mseitif, L.M. and Strom, M.S. 1988.** Evidence for plasmid-mediated toxin and bacteriocin production in *Clostridium botulinum* type G. *Appl. Environ. Microbiol.* **54**(6):1405-1408.
- Eklund, M.W., Poysky, F.T., Reed, S.M. and Smith, C.A. 1971.** Bacteriophage and the toxigenicity of *Clostridium botulinum* type C. *Science.* **172**:480-482.

- Elmore, M.J., Hutson, R.A., Collins, M.D., Bodsworth, N.J., Whelan, S.M. and Minton, N.P.** 1995. Nucleotide sequence of the gene coding for proteolytic (group I) *Clostridium botulinum* type F neurotoxin: Genealogical comparison with other Clostridal neurotoxins. *System. Appl. Microbiol.* **18**:23-31.
- Evans, D.M., Williams, R.S., Shone, C.C., Hambleton, P., Melling, J. and Dolly, J.O.** 1986. Botulinum neurotoxin type B. Its purification, radioiodination and interaction with rat-brain synaptosomal membranes. *Eur. J. Biochem.* **154**:409-416.
- Facchino, F., Benfenati, F., Valtorta, F. and Luini, A.** 1993. Covalent modification of synapsin I by a tetanus toxin-activated transglutaminase. *J. Biol. Chem.* **268**:4588-4591.
- Fach, P., Gibert, M., Griffais, R., Guillou, J.P. and Popoff, M.R.** 1995. PCR and gene probe identification of botulinum neurotoxin A-, B-, E-, F- and G-producing *Clostridium* spp. and evaluation in food samples. *Appl. Environ. Microbiol.* **61**(1):389-392.
- Fairweather, N.F. and Lyness, V.A.** 1986. The complete nucleotide sequence of tetanus toxin. *Nucl. Acids. Res.* **14**(19):7809-7812.
- Farber, J.M. and Peterkin, P.I.** 1991. *Listeria monocytogenes*, a food borne pathogen. *Microbiol. Rev.* **55**(3):476-511
- Farrand, S.K., Hwang, I. And Cook, D.M.** 1996. The *tra* region of the Nopaline-type Ti plasmid is a chimera with elements related to the transfer systems of RSF1010, RP4 and F. *J.Bacteriol.* **178**(14):4233-4247.
- Faulkner, J.D.B.** 1992. A novel series of expression vectors for use in the yeast *Saccharomyces cerevisiae*. Ph.D Thesis, University of Kent
- Feinstein, S.I. and Altman, S.** 1978. Context effects on nonsense codon suppression in *Escherichia coli*. *Genetics.* **88**:201-219.

- Ferreira, J.L. and Hamdy, M.K.** 1995. Detection of botulinal toxin genes: types A and E or B and F using the multiplex polymerase chain reaction. *J. Rapid methods and automation in Microbiol.* **3**:177-183.
- Ferreira, J.L., Hamdy, M.F., Zapatka, F.A. and Herbert, W.O.** 1981. Immunodiffusion method for detection of type A *Clostridium botulinum*. *Appl. Environ. Microbiol.* **42**(6):1057-1061.
- Ferreira, J.L., Hamdy, M.K., Zapatka, F.A. and McCay, S.G.** 1987. Plasmids of *Clostridium botulinum* type A, B, E, F and other clostridial species. *J. Food Safety.* **8**:267-277.
- Finlay, B.B. and Falkow, S.** 1997. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* **61**(2):136-169.
- Finn, C. W. Jr., Silver, R.P., Habig, W.H. and Hardegree, M.C.** 1984. The structural gene for tetanus neurotoxin is on a plasmid. *Science.* **224**:881-884.
- Firshein, W. and Kim, P.** 1997. Plasmid replication and partition in *Escherichia coli*: is the cell membrane the key? *Mol. Microbiol.* **23**(1):1-10.
- Fischer, R.J., Helms, J. and Durre, P.** 1993. Cloning, sequencing and molecular analysis of the *sol* operon of *Clostridium acetobutylicum*, a chromosomal locus involved in solventogenesis. *J. Bacteriol.* **175**(21):6959-6969.
- Fox, M.E., Lemmon, M.J., Mauchline, M.L., Davis, T.O., Giaccia, A.J., Minton, N.P. and Brown, J.M.** 1996. Anaerobic bacteria as a delivery system for cancer gene therapy: *in vitro* activation of 5-fluorocytosine genetically engineered *Clostridia*. *Gene Therapy.* **3**:173-178.

- Franciosa, G., Ferreira, J.L. and Hatheway, C.L.** 1994. Detection of type A, B and E botulism neurotoxin genes in *Clostridium botulinum* and other *Clostridium* species by PCR: evidence of unexpressed type B toxin genes in type A toxigenic organisms. *J. Clin. Microbiol.* **32**(8):1911-1917.
- Francis, C.L., Ryan, T.A., Jones, B.D., Smith, S.J. and Falkow, S.** 1993. Ruffles induced by *Salmonella* and other stimuli direct macropinocytosis of bacteria. *Nature.* **264**:639-642.
- Fujii, N., Kimura, K., Murakami, T., Indoh, T., Yashiki, T., Tsuzuki, K., Yokosawa, N. and Oguma, K.** The nucleotide and deduced amino acid sequences of *EcoRI* fragment containing the 5'-terminal region of *Clostridium botulinum* type E toxin gene cloned from Mashike, Iwanai and Otaru strains. *Microbiol. Immunol.* **34**(12)1041-1047.
- Fujii, N., Oguma, K., Yokosawa, N., Kimura, K. and Tsuzuki, K.** 1988. Characterisation of bacteriophage nucleic acids obtained from *Clostridium botulinum* types C and D. *Appl. Environ. Microbiol.* **54**(1):69-73.
- Fujinaga, Y., Inoue, K., Shimazaki, S., Tomochika, K., Tsuzuki, K., Fujii, N., Watanabe, T., Ohyama, Y., Takeshi, K., Inoue, K. and Oguma, K.** 1994. Molecular construction of *Clostridium botulinum* type C progenitor toxin and its gene organisation. *Biochem. Biophys. Res. Comm.* **205**(2):1291-1298.
- Fujinaga, Y., Inoue, K., Watanabe, S., Yokota, K., Hirai, Y., Nagamachi, E. and Oguma, K.** 1997. The haemagglutinin of *Clostridium botulinum* type C progenitor toxin plays an essential role in binding of toxin to the epithelial cells of guinea pig small intestine, leading to the efficient absorption of the toxin. *Microbiol.* **143**:3841-3847.

- Fujinaga, Y., Takeshi, K., Inoue, K., Fujita, R., Tohru, O., Moriishi, K. and Oguma, K.** 1995. Type A and B neurotoxin genes in a *Clostridium botulinum* type AB strain. *Biochem. Biophys. Res. Comm.* **213**(3):737-745.
- Fujita, R., Fujinaga, Y., Inoue, K., Nakajima, H., Kumon, H. and Oguma, K.** 1995. Molecular characterisation of two forms of nontoxic-nonhemagglutinin components of *Clostridium botulinum* type A progenitor toxins. *FEBS Letts.* **376**:41-44.
- Garnier, T. and Cole, S.T.** 1988a. Studies of UV-inducible promoters from *Clostridium perfringens* *in vivo* and *in vitro*. *Mol. Microbiol.* **2**(5):607-614.
- Garnier, T. and Cole, S.T.** 1988b. Complete nucleotide sequence and genetic organisation of the bacteriocinogenic plasmid, pIP404, from *Clostridium perfringens*. *Plasmid.* **19**:151-160.
- Garnier, T. and Cole, S.T.** 1988c. Identification and molecular genetic analysis of replication functions of the bacteriocinogenic plasmid pIP404 from *Clostridium perfringens*. *Plasmid* **19**:151-160.
- Garnier, T., Canard, B. and Cole, S.T.** 1991. Cloning, mapping and molecular characterisation of the rRNA operons of *Clostridium perfringens*. *J. Bacteriol.* **173**:5431-5438.
- Gennaro, M.L., Kornblum, J. and Novick, R.P.** 1987. A site-specific recombination function in *Staphylococcus aureus* plasmids. *J. Bacteriol.* **169**(6):2601-2610.
- Geppert, M., Goda, Y., Hammer, R.E., Li, C., Rosahl, T.W., Stevens, C.F. and Sudhof, T.C.** 1994. Synaptotagmin I: a major Ca^{2+} sensor for transmitter release at a central synapse. *Cell.* **79**:717-727.

- Gerscher, U. and Durre, P.** 1990. Cloning, sequencing and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J. Bacteriol. 172(12):6907-6918.
- Gibbs, C.P., Reimann, B-Y., Schultz, E., Kaufmann, A., Haas, R. and Meyer, T.F.** 1989. Reassortment of pilin genes in *Neisseria gonorrhoeae* occurs by two distinct mechanisms. Nature. 338:651-652.
- Gibson, T.J.** 1984. Studies on the Epstein-Barr virus genome. Ph.D. thesis, Cambridge University, England.
- Gimenez, D.F.** 1984. *Clostridium botulinum* subtype Ba. Zbl. Bakt. Hyg. A. 257:68-72.
- Gimenez, D.F. and Ciccarelli, A.S.** 1970. Another type of *Clostridium botulinum*. Zentralbl. Bakteriologie. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. Reihe A. 215:221-224.
- Gimenez, D.F. and Ciccarelli, A.S.** 1970. Studies on strain 84 of *Clostridium botulinum*. Zentralbl. Bakteriologie. Parasitenkd. Infektionskr. Hyg. Abt. 1 Orig. Reihe A. 215:212-220.
- Gimenez, D.F. and Ciccarelli, A.S.** 1972. Variaciones antigenicas en toxinas botulinicas del tipo F. Ensayo de definiciones para la tipificacion serologica y clasificacion de *Clostridium botulinum*. Medicina. 32:596-606.
- Girbal, L. and Soucaille, P.** 1998. Regulation of solvent production in *Clostridium acetobutylicum*. TIBTECH. 16:11-16.
- Goda, S.K. and Minton, N.P.** 1995. A simple procedure for gel electrophoresis and Northern blotting of RNA. Nucleic acids Res. 23(16):3357-3358.
- Goodnough, M.C., Hammer, B., Sugiyama, H. and Johnson, E.A.** 1993. Colony immunoblot assay of botulinum toxin. Appl. Environ. Microbiol. 59(7):2339-2342.

- Gozalbo, D and Hohmann, S.** 1989. The naturally occurring silent invertase structural gene *suc2^o* contains an amber stop codon that is occasionally read through. *Mol. Gen. Genet.* **216**:511-516.
- Graves, M.C., Mullenbach, G.T. and Rabinowitz, J.C.** 1985. Cloning and nucleotide sequence determination of the *Clostridium pasteurianum* ferredoxin gene. *Proc. Natl. Acad. Sci. USA.* **82**:1653-1657.
- Gross, C.A., Lonetto, M. and Losick, R.** 1992. Bacterial sigma factors. In: *Transcriptional regulation*, K. Yamamoto and S. McKnight, eds. P. 129-176. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
- Grunstein, M. and Hogness, D.S.** 1975. Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA.* **72**(10):3961-3965
- Gruss, A. and Ehrlich, D.S.** 1989. The family of highly interrelated single-stranded deoxyribonucleic acid plasmids. *Microbiol. Rev.* **53**(2):231-241.
- Hall, J. D., McCroskey, L. M., Pincomb, B. J. and Hatheway, C. L.** 1985. Isolation of an organism resembling *Clostridium barati* which produces type F botulinal toxin from an infant with botulism. *J. Clin. Microbiol.* **21**:654-655.
- Halpern, J.L. and Loftus, A.** 1993. Characterisation of the receptor-binding domain of tetanus toxin. *J. Biol. Chem.* **268**(15):11188-11192.
- Halpern, J.L., Smith, L.A., Seamon, K.B., Groover, K.A. and Habig, W.H.** 1989. Sequence homology between tetanus and botulinum toxins detected by an antipeptide antibody. *Infect. Immun.* **57**(1):18-22.
- Hammond, G.A. and Johnson, J.L.** 1995. The toxigenic element of *Clostridium difficile* strain VPI 10463. *Microbial Path.* **19**:203-213.

- Hammond, G.A., Lyster, D.M. and Johnson, J.L.** 1997. Transcriptional analysis of the toxigenic element of *Clostridium difficile*. Microbial Path. 22:143-154.
- Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557-580.
- Hardy, K. G.** 1988. Plasmids : A practical approach. p. 1-176. Ed Hardy, K.G. IRL Press, Oxford, U.K.
- Harris, A.J. and Miledi, R.** 1971. The effect of type D botulinum toxin on frog neuromuscular junctions. J. Physiol. 217:497-515.
- Hatheway, C. L.** 1995. Botulism: The present status of the disease. In: Current Topics in Microbiology and Immunology (195) P.55-75. Ed Montecucco, C. Springer-Verlag Berlin Germany.
- Hatheway, C. L., McCroskey, L. M., Lombard, G. L. and Dowell, V. R. Jr.** 1981. Atypical toxin variant of *Clostridium botulinum* type B associated with infant botulism. J. Clin. Microbiol. 14:607-611.
- Hatheway, C.L. and McCroskey, L.M.** 1989. Unusual neurotoxigenic clostridia recovered from human fecal specimens in the investigation of botulism, p.477-481. In: Proceedings of the 5th International Symposium on Microbiol Ecology: recent advances in microbial ecology. Eds. T. Hattori, Y.Ishida, Y. Maruyama, R.Y.Morta and A.uchida . Japan Scientific Societies Press, Tokyo.
- Hauser, D., Eklund, M.W, Boquet, P. and Popoff, M.R.** 1994. Organization of the botulinum neurotoxin C1 gene and its associated non-toxic protein genes in *Clostridium botulinum* C 468. Mol. Gen. Genet. 243:631-640.

- Hauser, D., Eklund, M.W., Kuazono, H., Binz, T., Niemann, H., Gill, D.M., Boquet, P. and Popoff, M.R.** 1990. Nucleotide sequence of the *Clostridium botulinum* C1 neurotoxin. Nucleic Acids Res. 18:4924.
- Hauser, D., Gibert, M., Boquet, P. and Popoff, M.R.** 1992. Plasmid localisation of a type E botulinal neurotoxin gene homologue in toxigenic *Clostridium butyricum* strains and absence of this gene in non-toxigenic *C. butyricum* strains. FEMS Microbiol. Lett. 99:251-256.
- Hawley, D.K. and McClure, W.R.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucl. Acids Res. 11:2237-2255.
- Heefner, D.L., Squires, C.H., Evans, R.J., Kopp, B.J. and Yarus, M.J.** 1984. Transformation of *Clostridium perfringens*. J. Bacteriol. 159(2):460-464.
- Higgins, C.F., Peltz, S.W. and Jacobson, A.** 1992. Turnover of mRNA in prokaryotes and lower eukaryotes. Curr. Opin. Genet. Dev. 2:739-747.
- Higgins, D.E. and DiRita, V.J.** 1996. Genetic analysis of the interaction between *Vibrio cholerae* transcription activator ToxR and *toxT* promoter DNA. J. Bacteriol. 178(4):1080-1087.
- Higgins, D. G. and Sharp, P.M.** 1988. CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene 73: 237-244
- Hoch, D.H., Romera-Mira, M., Ehrlich, B.E., Finkelstein, A., DasGupta, B.R. and Simpson, L.L.** 1985. Channels formed by botulinum, tetanus and diphtheria toxins in planar lipid bilayers: relevance to translocation of proteins across membranes. Proc. Natl. Acad. Sci. USA. 82:1692-1696.

- Holmes, D.S. and Quigley, M.** 1981. A rapid boiling method for the preparation of bacterial plasmids. *Analyt. Biochem.* **114**:193-197.
- Hu, G.** 1993. DNA polymerase-catalysed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. *DNA and Cell Biol.* **12**(8):763-770
- Hua, J. and Patton, J.T.** 1994. The carboxyl-half of the rotavirus nonstructural protein NS53 (NSP1) is not required for virus replication. *Virol.* **198**:567-576.
- Huang, X., Decatur, A., Sorokin, A. and Helmann, J.D.** 1997. The *Bacillus subtilis* σ^x protein is an extracytoplasmic function σ factor contributing to survival at high temperature. *J. Bacteriol.* **179**(9):2915-2921.
- Hue, K. K., Cohen, S. D. and Bechhofer, D. H.** 1995. A polypurine sequence that acts as a 5' mRNA stabiliser in *Bacillus subtilis*. *J. Bacteriol.* **177**(12):3465-3471.
- Hultgren, S.J., Normark, S. and Abraham, S.N.** 1991. Chaperone-assisted assembly and molecular architecture of adhesive pili. *Ann. Rev. Microbiol.* **45**:383-415
- Hutson, R.A., Collins, M.D., East, A.K. and Thompson, D.E.** 1994. Nucleotide sequence of the gene coding for non-proteolytic *Clostridium botulinum* type B neurotoxin: Comparison with other Clostridial neurotoxins. *Current Microbiol.* **28**:101-110.
- Hutson, R.A., Zhou, Y., Collins, M.D., Johnson, E.A., Hatheway, C.L. and Sugiyama, H.** 1996. Genetic characterisation of the *Clostridium botulinum* type A containing silent type B neurotoxin gene sequences. *J. Biol. Chem.* **271**(18):10786-10792.
- Inoue, K. and Iida, H.** 1968. Bacteriophages of *Clostridium botulinum*. *J. Virol.* **2**(5):537-540.
- Inoue, K. and Iida, H.** 1970. Conversion of toxigenicity in *Clostridium botulinum* type C. *Japan J. Microbiol.* **14**(1):87-89.

- Inoue, K., Fujinaga, Y., Watanabe, T., Ohyama, T., Takeshi, K., Moriishi, K., Nakajima, H., Inoue, K. and Oguma, K.** 1996. Molecular composition of *Clostridium botulinum* type A progenitor toxins. *Infect. Immun.* **64**(5):1589-1594.
- Inouye, S., Sunshine, M.G., Six, E.W. and Inouye, M.** 1991. Retronphage ϕ R73: An *E. coli* phage that contains a retroelement in a tRNA gene. *Science.* **252**:969-971
- Isberg, R.R. and Leong, J.M.** 1990. Multiple β_1 chain integrins are receptors for invasins, a protein that promotes bacterial penetration into mammalian cells. *Cell.* **60**:861-871.
- Ish-Horowicz, D. and Burke, J.F.** 1981. Rapid and efficient cosmid cloning. *Nucl. Acid. Res.* **9**:2989-2998
- Izard, J.W. and Kendall, D.A.** 1994. Signal peptides: exquisitely designed transport promoters. *Mol. Microbiol.* **13**(5):765-773.
- Janniere, L., Bruand, C. and Ehrlich, D.S.** 1990. Structurally stable *Bacillus subtilis* cloning vectors. *Gene.* **87**:53-61.
- Johnson, E.A.** 1997. Extrachromosomal virulence determinants in the clostridia. In *The Clostridia: Molecular Biology and Pathogenesis*. P.35-48 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Johnston, C., Pegues, D.A., Hueck, C.J., Lee, C.A. and Miller, S.I.** 1996. Transcriptional activation of *Salmonella typhimurium* invasion genes by a member of the phosphorylated response-regulator superfamily. *Mol. Microbiol.* **22**(4):715-727.
- Jones, D. T. and Woods, D. R.** 1989. Acetone-butanol fermentation revisited. *Microbiol. Rev.* **50**:484-524.
- Josson, K., Soetaert, P., Michiels, F., Joos, H. and Mahillon, J.** 1990. *Lactobacillus higaradii* plasmid pLAB1000 consists of two functional cassettes commonly found in other Gram-positive organisms. *J. Bacteriol.* **172**(6):3089-3099.

- Kahn, D. and Ditta, G.** 1991. Modular structure of FixJ: homology of the transcriptional activator domain with the -35 binding domain of sigma factors. *Mol. Microbiol.* **5**(4):987-997.
- Kalmbach, E.R.** 1968. Type C botulism among wild birds-A historical sketch. Bureau of Sport Fisheries and Wildlife, Special Scientific Report, Wildlife. **110**. Washington, D.C.
- Kapur, V., Kanjilal, S., Hamrick, M.R., Li, L-L., Whittam, T.S., Sawyer, S.A. and Musser, J.M.** 1995. Molecular population genetic analysis of the streptokinase gene of *Streptococcus pyogenes*: mosaic alleles generated by recombination. *Mol. Microbiol.* **16**(3):509-519.
- Kawata, T. and Takumi, K.** 1971. Autolytic enzyme system of *Clostridium botulinum*. *Japan J. Microbiol.* **15**(1):1-10.
- Khan, S.A.** 1996. Mechanism of replication and copy number control of plasmids in Gram-positive bacteria. In *Genetic engineering*, Vol. **18**, P. 183-201. Ed. Setlow, J.K. Plenum Press, NY.
- Kiewiet, R., Kok, J., Seegers, J.F.M.L., Venema, G. and Bron, S.** 1993. The mode of replication is a major factor in segregational plasmid instability in *Lactococcus lactis*. *Appl. Environ. Microbiol.* **59**(2):358-364.
- Kimura, K., Fujii, N., Tsuzki, K., Murakami, T., Indoh, T., Yokosawa, K., Takeshi, K., Syuto, B. and Oguma, K.** 1990. The complete nucleotide sequence of the gene coding for botulinum type C₁ toxin in the C-ST phage genome. *Biochem. Biophys. Res. Comm.* **171**(3):1304-1311.
- Kindler, S.H., Mager, J. and Grossowicz, N.** 1956. Toxin production by *Clostridium parabotulinum* type A. *J. Gen. Microbiol.* **15**:394-403.

- Kinouchi, T., Takumi, K. and Kawata, T.** 1981. Characterisation of two inducible bacteriophages, $\alpha 1$ and $\alpha 2$, isolated from *Clostridium botulinum* type A 190L and their deoxyribonucleic acids. *Microbiol. Immunol.* **25**(9):915-927.
- Kitamura, M., Iwamori, M. and Nagai, Y.** 1980. Interaction between *Clostridium botulinum* neurotoxin and gangliosides. *Biochim. Biophys. Acta.* **628**:328-335.
- Kleerebezem, M., Quadri, L.E.N., Kuipers, O.P. and de Vos, W.M.** 1997. Quorum sensing by peptide pheromones and two component signal-transduction systems in Gram-positive bacteria. *Mol. Microbiol.* **24**(5):895-904.
- Klier, A., Msadek, T. and Rapoport, G.** 1992. Positive regulation in the Gram-positive bacterium: *Bacillus subtilis*. *Ann. Rev. Microbiol.* **46**:429-459.
- Kok, J., Van der Vossen, J.M.B.M. and Venema, G.** 1984. Construction of plasmid cloning vectors for lactic streptococci which also replicate in *Bacillus subtilis* and *Escherichia coli*. *Appl. Environ. Microbiol.* **48**:726-731.
- Kokai-Kun, J.F. and McClane, B.A.** 1997. Determination of functional regions of *Clostridium perfringens* enterotoxin through deletion analysis. *Clin. Infect. Dis.* **S165-S167**.
- Kramer, M.G., Kahn, S.A. and Espinosa, M.** 1997. Plasmid rolling circle replication: identification of the RNA polymerase-directed primer RNA and requirement for DNA polymerase I for lagging strand synthesis. *EMBO J.* **16**(18):5784-5795.
- Kriegelstein, K.G., DasGupta, B.R. and Henschen, A.H.** 1994. Covalent structure of botulinum neurotoxin type A: location of sulfhydryl groups and disulfide bridges and identification of C-termini of light and heavy chains. *J. Prot. Chem.* **13**(1): 49-57.
- Kubota, T, Shirakawa, S., Kozaki, S., Isogai, E., Isogai, H., Kimura, K. and Fujii, N.** 1996. Mosaic type of the nontoxic-nonhemagglutinin component gene in *Clostridium*

botulinum type A strain isolated from infant botulism in Japan. Biochem. Biophys. Res. Comm. 224:843-848.

Kurazono, H., Mochida, S., Binz, T., Eisel, U., Quanz, M., Grebenstein, O., Wernars, K., Poulain, B., Tuac, L. and Niemann, H. 1992. Minimal essential domains specifying toxicity of the light chains of tetanus toxin and botulinum neurotoxin type A. J. Biol. Chem. 267(21):14721-14729.

Kurose, N., Miyazaki, T., Kakimoto, T., Yagyu, J., Uchida, M., Obayashi, A. and Murooka, Y. 1989. Isolation of plasmids from thermophilic clostridia and construction of shuttle vectors in *Escherichia coli* and cellulolytic clostridia. J. Ferment. Bioengin. 68(5):371-374.

Laird, W.J., Aaronson, W., Silver, R.P., Habig, W.H., Hardegree, M.C. 1980. Plasmid-associated toxigenicity in *Clostridium tetani*. J. Infect. Dis. 142:623.

Lebeda, F.J. and Olson, M.A. 1994. Secondary structural predictions for the Clostridial neurotoxins. Protein: Structure, Function and Genetics. 20:293-300.

Lebeda, F.J. and Olson, M.A. 1995. Structural predictions of the channel-forming region of botulinum neurotoxin heavy chain. Toxicon. 33(4):559-567.

Leblanc, D.J. and Lee, L.N. 1984. Physical and genetic analysis of Streptococcal plasmid pAM β 1 and cloning of its replication region. J. Bacteriol. 157(2):445-453

Lee, C.A. 1996. Pathogenicity islands and the evolution of bacterial pathogens. Infect. Agents and Dis. 5:1-7.

Lee, C-K., Durre, P., Hippe, H. and Gottschalk, G. 1987. Screening for plasmids in the genus *Clostridium*. Arch. Microbiol. 148:107-114.

- Lemmel, S.A.** 1985. Mutagenesis in *Clostridium acetobutylicum*. *Biotech. Lett.* 7(10):711-716.
- Lin, W-J.** 1992. The development of genetic methods to study group I *Clostridium botulinum*. Ph.D. thesis, University of Wisconsin, Madison.
- Lin, W-J. and Johnson, E.A.** 1991. Transposon Tn916 mutagenesis in *Clostridium botulinum*. *Appl. Environ. Microbiol.* 57(10):2946-2950.
- Lin, W-J. and Johnson, E.A.** 1995. Genome analysis of *Clostridium botulinum* type A by pulsed-field gel electrophoresis. *Appl. Environ. Microbiol.* 61(12):4441-4447.
- Lin, Y-L. and Blasheck, H. P.** 1984. Transformation of heat-treated *Clostridium acetobutylicum* protoplasts with pUB110 plasmid DNA. *Appl. Environ. Microbiol.* 48(4):737-742.
- London, E.** 1992. How bacterial protein toxins enter cells: the role of partial unfolding in membrane translocation. *Mol. Microbiol.* 6(22):3277-3282.
- Lonetto, M., Gribskov, M. and Gross, C.A.** 1992. The σ^{70} family: Sequence conservation and evolutionary relationships. *J. Bacteriol.* 174(12):3843-3849.
- Lonetto, M.A., Brown, K.L., Rudd, K.E. and Buttner, M.J.** 1994. Analysis of the *Streptomyces coelicolor sigE* gene reveals the existence of a subfamily of eubacterial RNA polymerase σ factors involved in the regulation of extracytoplasmic functions. *Proc. Natl. Acad. Sci. USA.* 91:7573-7577.
- Lowenthal, J.P. and Lamanna.** 1952. Characterisation of botulinum hemagglutination. *Am. J. Hyg.* 57:46-59.
- Lyras, D. and Rood, J.I.** 1997. Transposable genetic elements and antibiotic resistance determinants from *Clostridium perfringens* and *Clostridium difficile*. In *The Clostridia*:

- Molecular Biology and Pathogenesis. p.73-92 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Maisey, E.A., Wadsworth, J.D.F., Poulain, B., Shone, C.C., Melling, J., Gibbs, P., Tauc, L. and Dolly, J.O.** 1988. Involvement of the constituent chains of botulinum neurotoxins A and B in the blockade of neurotransmitter release. *Eur. J. Biochem.* **177**:683-691.
- McCroskey, L.M., Hatheway, L.F., Pasolini, B. and Aureli, P.** 1986. Characterisation of an organism that produces type E botulinal toxin but which resembles *Clostridium butyricum* from feces of an infant with type E botulism. *J. Clin. Microbiol.* **23**:201-202.
- McDonell, M.W., Simon, M.N. and Studier, F.W.** 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. *J. Mol. Biol.* **110**:119-146.
- McKinney, M.W., Levett, P.N. and Haylock, R.W.** 1993. Cloning of a DNA sequence unique to *Clostridium botulinum* group I by selective hybridisation. *J. Clin. Microbiol.* **31**(7):1845-1849.
- Mermelstein, L. D., Welker, N. E., Bennett, G. N. and Papoutsakis, E. T.** 1992. Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. *Biotechnol.* **10**:190-195.
- Mermelstein, L.D. and Papoutsakis, E.T.** 1993. *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage ϕ 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* **59**(4):1077-1081.
- Miller, J.** 1994. Bacterial transformation by electroporation. *Meth. Enzymol.* **235**:375-385.

- Minton, N. and Morris, J.G.** 1981. Isolation and partial characterisation of three cryptic plasmids from strains of *Clostridium butyricum*. J. Gen. Micro. 127:325-331.
- Minton, N.P., Brehm, J.K., Swinfield, T-J., Whelan, S.M., Mauchline, M.L., Bodsworth, N. and Oultram, J.D.** 1993. Clostridial cloning vectors. In: The Clostridia and biotechnology, P. 119-150. Ed. Woods, D.R. Butterworth-Heinemann. U.S.A.
- Miyasaki, S., Iwasaki, M. and Sakaguchi, G.** 1977. *Clostridium botulinum* type D toxin; purification, molecular structure and some immunological properties. Infect. Immun. 17:395-401.
- Moberg, L. J. and Sugiyama, H.** 1979. Microbial ecological basis of infant botulism as studied with germ free mice. Infect. Immun. 25:653-657
- Mochida, S., Poulain, B., Weller, U., Habermann, E. and Tauc, L.** 1989. Light chain of tetanus toxin intracellularly inhibits acetylcholine release at neuro-neuronal synapses and its internalisation mediated by heavy chain. FEBS Lett. 253(1,2): 47-51.
- Moncrief, J.S., Barroso, L.A. and Wilkins, T.D.** 1997. Positive regulation of *Clostridium difficile* toxins. Infect. Immun. 65(3):1105-1108.
- Moncrief, J.S., Lyerly, D.M. and Wilkins, T.D.** 1997. Molecular biology of *Clostridium difficile* toxins. In The Clostridia: Molecular Biology and Pathogenesis. P.369-392 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Montal, M.S., Blewitt, R., Tomich, J.M. and Montal, M.** 1992. Identification of an ion channel-forming motif in the primary structure of tetanus and botulinum neurotoxins. FEBS lett. 313(1):12-18.
- Montecucco, C.** 1986. How do tetanus and botulinum toxins bind to neuronal membranes? TIBS. 11:314:317.

- Montecucco, C. and Schiavo, G.** 1993. Tetanus and botulism neurotoxins: a new group of zinc proteases. *TIBS*. 18:324-327.
- Montecucco, C., Schiavo, G., Tugnoli, V. and de Grandis, D.** 1996. Botulinum neurotoxins: mechanism of action and therapeutic applications. *Mol. Med. Today*. 418-424
- Moran, C. P. Jr.** 1993. RNA Polymerase and transcription factors. In *Bacillus subtilis* and other Gram-positive organisms. P.653-667 Eds. Sonenshein, A.L., Hoch, J.A. and Losick, R.. A.S.M. Washington D.C.
- Moriishi, K, Koura, M., Fujii, N., Fujiaga, Y., Inoue, K., Syuto, B. and Oguma, K.** 1996. Molecular cloning of the gene encoding the mosaic neurotoxin, composed of parts of botulinum neurotoxin types C1 and D and PCR detection of this gene from *Clostridium botulinum* type C organisms. *Appl. Environ. Microbiol.* 62(2):662-667.
- Moriishi, K., Koura, M., Abe, N., Fujii, N., Fujiaga, Y., Inoue, K. and Oguma, K.** 1996. Mosaic structures of neurotoxins produced from *Clostridium botulinum* types C and D organisms. *Biochim. Biophys. Acta*. 1307:123-126.
- Morris, J.G.** 1993. History and future potential of the Clostridia in biotechnology. In *The Clostridia and Biotechnology*. P.1-23. Ed. Woods, D.R. Butterworth-Heinemann. U.S.A.
- Morschhauser, J., Vetter, V., Emody, L. and Hacker, J.** 1994. Adhesin regulatory genes within large, unstable DNA regions of pathogenic *Escherichia coli*: cross-talk between different adhesin gene clusters. *Mol. Microbiol.* 11(3):555-566.
- Morton, T.M., Eaton, D.M., Johnston, L. and Archer, G.L.** 1993. DNA sequence and units of transcription of the conjugative transfer gene complex (*trs*) of *Staphylococcus aureus* plasmid pGO1. *J. Bacteriol.* 175(14):4436-4447.

- Neimann, H., Blasi, J. and Jahn, R.** 1994. Clostridial neurotoxins: New tools for dissecting exocytosis. *Trends Cell. Biol.* 4:179-185.
- Nishiki, T., Kamata, Y., Nemoto, Y., Omori, A., Ito, T., Takahashi, M. and Kosaki, S.** 1994. Identification of protein receptor for *Clostridium botulinum* type B neurotoxin in rat brain synaptosomes. *J. Biol. Chem.* 269:10498-10503.
- Nobes, C.D. and Hall, A.** 1995. Rho, Rac, and Cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibres, lamellipodia, and filopodia. *Cell.* 81:53-62.
- Noirot-Gros, M. F., Bidnenko, V. and Ehrlich, S. D.** 1994. Active site of the replication protein of the rolling circle plasmid pC194. *EMBO J.* 13(18):4412-4420.
- Novick, R.P.** 1989. Staphylococcal plasmids and their replication. *Ann. Rev. Microbiol.* 43:537-565.
- Nukina, M., Mochida, Y., Sakaguchi, S. and Sakaguchi, G.** 1991. Difficulties of molecular dissociation of *Clostridium botulinum* type G progenitor toxin. *FEMS Microbiol. Letts.* 79:165-170.
- Ochanda, J.O., Syuto, B., Oguma, K., Iida, H. and Kubo, S.** 1984. Comparison of antigenicity of toxins produced by *Clostridium botulinum* type C and D strains. *Appl. Environ. Microbiol.* 47(6):1319-1322.
- Ogasawara, J., Kamata, Y., Sakaguchi, G. and Kozaki, S.** 1991. Properties of a protease-sensitive acceptor component in mouse brain synaptosomes for *Clostridium botulinum* type B neurotoxin. *FEMS Microbiol. Lett.* 79:351-356.

- Ogert, R.A., Brown, J.E., Singh, B. R., Shriver-Lake, L.S. and Ligler, F.S. 1992. Detection of *Clostridium botulinum* toxin A using a fiber optic-based biosensor. *Analyt. Biochem.* 205:306-312.
- Oguma, K., Iida, H. and Shiozaki, M. 1976. Phage conversion to hemagglutinin production in *Clostridium botulinum* types C and D. *Infect. Immun.* 14(3):597-602.
- Ohyama, T., Watanabe, T., Fujinaga, Y., Inoue, K., Suagawa, H., Fujii, N., Inoue, K. and Oguma, K. 1995. Characterisation of nontoxic-nonhemagglutinin component of the two types of progenitor toxin (M and L) produced by *Clostridium botulinum* type D CB-16. *Microbiol. Immunol.* 39(7):457-465.
- Okabe, A. and Cole, S. T. 1997. Extracellular enzymes from *Clostridium perfringens* and *Clostridium histolyticum* that damage connective tissue. In *The Clostridia: Molecular Biology and Pathogenesis*. P.411-422 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Olek, A., Oswald, J. and Walter, J. 1996. A modified method for bisulphite based cytosine methylation analysis. *Nucl. Acids. Res.* 24(24):5064-5066.
- Oultram, J.D., Loughlin, M., Swinfield, T-J., Brehm, J.K., Thompson, D.E. and Minton, N.P. 1988. Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. *FEMS Microbiol. Lett.* 56:83-88.
- Oultram, J.D., Peck, H., Brehm, J.K., Thompson, D.E., Swinfield, T.J. and Minton, N.P. 1988. Introduction of genes for leucine biosynthesis from *Clostridium pasteurianum* into *C. acetobutylicum* by cointegrate conjugal transfer. *Mol. Gen. Genet.* 214(1):177-179.
- Pabo, C.O. and Sauer, R.T. 1992. Transcription factors: structural families and principles of DNA recognition. *Ann. Rev. Biochem.* 61:1053-1095.

- Paillart, J-C., Skripkin, E., Ehresmann, B., Ehresmann, C. and Marquet, R.** 1996. A loop-loop 'kissing' complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. USA.* **93**:5572-5577.
- Palmer, B.R. and Marinas, M.G.** 1994. *Gene.* **143**:1-12
- Patterson-Curtis, S.I. and Johnson, E.A.** 1989. Regulation of neurotoxin and protease formation in *Clostridium botulinum* Okra B and Hall A by arginine. *Appl. Environ. Microbiol.* **55**(6):1544-1548.
- Patti, J.M., Allen, B.L., McGavin, M.J. and Hook, M.** 1994. MSCRAMM-mediated adherence of microorganisms to host tissues. *Ann. Rev. Microbiol.* **48**:585-617
- Phillips-Jones, M.K.** 1995. Introduction of recombinant DNA into *Clostridium* spp. In *Methods in molecular biology, Vol 47: Electroporation protocols for microorganisms.* Ed. J.A. Nickoloff. Humana Press Inc., Totwa, NJ.
- Pich, A. and Bahl, H.** 1991. Purification and characterisation of the DNA-dependent RNA polymerase from *Clostridium acetobutylicum*. *J. Bacteriol.* **173**:2120-2124.
- Popoff, M.R. and Truffaut, N.** 1985. Survey of plasmids in *Clostridium butyricum* and *Clostridium beijerinckii* strains from different origins and different phenotypes. *Curr. Microbiol.* **12**:151-156.
- Poulain, B., Mochida, S., Wadsworth, J.D.F., Weller, U., Habermann, E., Dolly, J.O. and Tauc, L.** 1990. Inhibition of neurotransmitter release by botulinum neurotoxins and tetanus toxin at Aplysia synapses: role of the constituent chains. *J. Physiol. (Paris).* **84**:247-261.
- Poulet, S., Hauser, D., Quanz, M., Niemann, H. and Popoff, M.R.** 1992. Sequences of the botulinal neurotoxin E derived from *Clostridium botulinum* type E (strain Beluga) and

- Clostridium butyricum* (strains ATCC 43181 and ATCC 43755). Biochem. Biophys. Res. Comm. 183(1):107-113
- Poumeyrol, M., Billon, J., Delille, F., Haas, C., Marmonier, A. and Sebald, M. 1983. Intoxification botulique mortelle due a une souche de *Clostridium botulinum* de type AB. Med. Malad. Infect. 13:750-754.
- Priebe, S.D. and Lacks, S.A. 1989. Region of the streptococcal plasmid pMV158 required for conjugative mobilisation. J. Bacteriol. 171:4778-4784.
- Projan, S.J., Moghazeh, S. and Novick, R.P. 1988. Nucleotide sequence of pS194, a streptomycin-resistance plasmid from *Staphylococcus aureus*. Nucl. Acid. Res. 16(5):2179-2187.
- Rasooly, A. and Rasooly, R.S. 1997. How rolling circle plasmids control their copy number. TMS. 5(11):440-446.
- Reed, K.C. and Mann, D.A. 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucl. Acid. Res. 13(20):7207-7221
- Reysset, G. and Sebald, M. 1985. Conjugal transfer of plasmid-mediated antibiotic resistance from streptococci to *Clostridium acetobutylicum*. Ann. Inst. Pasteur/Microbiol. 136:275-282.
- Reysset, G., Hubert, J., Podvin, L. and Sebald, M. 1987. Protoplast formation and regeneration of *Clostridium acetobutylicum* strain N1-4080. J. Gen. Microbiol. 133:2595-2600.
- Reysset, G., Hubert, J., Podvin, L. and Sebald, M. 1988. Transfection and transformation of *Clostridium acetobutylicum* strain N1-4081 protoplasts. Biotech. Tech. 2(3):199-204.

- Rhen, M., Vaisanen-Rhen, V., Saraste, M. and Korhonen, T.K.** 1986. Organisation of genes expressing the blood-group-M-specific hemagglutinin of *Escherichia coli*: identification and nucleotide sequence of the M-agglutinin subunit gene. *Gene*. **49**:351-360.
- Robinson, J.P. and Hash, J.H.** 1982. A review of the molecular structure of tetanus toxin. *Mol. Cell Biochem.* **48**:33-44.
- Roggentin, P. and Schauer, R.** 1997. Clostridial sialidases. In *The Clostridia: Molecular Biology and Pathogenesis*. P.423-437 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Romeo, J.M., Esmon, B. and Zusman, D.R.** 1986. Nucleotide sequence of the myxobacterial hemagglutinin gene contains four homologous domains. *Proc. Natl. Acad. Sci. USA*. **83**:6332-6336.
- Rood, J.I.** 1997. Genetic analysis in *Clostridium perfringens*. In *The Clostridia: Molecular Biology and Pathogenesis*. P.65-72 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Rosqvist, R., Hakansson, S., Forsberg, A. and Wolf-Waltz, H.** 1995. Functional conservation of the secretion and translocation machinery for virulence proteins of yersiniae, salmonellae and shigellae. *EMBO J.* **14**(17):4187-4195.
- Roth, C., Bringaud, F., Layden, R.E., Baltz, T. and Eisen, H.** 1989. Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc. Natl. Acad. Sci. USA*. **86**:9375-9379.
- Russell, R.R.B., Aduse-Opoku, J., Sutcliffe, I.C., Tao, L. and Ferretti, J.J.** 1992. A binding protein-dependent transport system in *Streptococcus mutans* responsible for multiple sugar metabolism. *J. Biol. Chem.* **267**(7):4631-4637.

- Russo, F.D. and Silhavy, T.J.** 1993. The essential tension: opposed reactions in bacterial two-component regulatory systems. *TIMS* 1(8):306-310.
- Ryoji, M., Hsia, K. and Kaji, A.** 1983. Read-through translation. *TIBS*. 8:88-89.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B and Erlich, H.A.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. 239:487-491.
- Sakaguchi, G., Ohishi, I. and Kozaki, S.** 1989. Botulism-Structure and chemistry of botulinum. P191-219. In *Microbial Toxins*. Eds. Kadis, S. and Montie, T.C. AGL
- Sakai, T., Sasakawa, C. and Yoshikawa, M.** 1988. Expression of four virulence antigens of *Shigella flexneri* is positively regulated at the transcriptional level by the 30 kilo Dalton *virF* protein. *Mol. Microbiol.* 2(5): 589-597.
- Salmond, G.P.C., Bycroft, B.W., Stewart, G.S.A.B. and Williams, P.** 1995. The bacterial 'enigma':cracking the code of cell-cell communication. *Mol. Microbiol.* 16(4):615-624.
- Sambrook, J., Fritsch, E.F. and Maniatis, T.** 1989. *Molecular cloning: A laboratory manual*, second ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., USA.
- Sandler, P. and Weisblum, B.** 1989. Erythromycin-induced ribosome stall in the *ermA* leader: a barricade to the 5'-to-3' nucleolytic cleavage of the *ermA* transcript. *J. Bacteriol.* 171:6680-6688.
- Sandros, J. and Tuomanen, E.** 1993. Attachment factors of *Bordetella pertussis*: mimicry of eucaryotic cell recognition molecules. *Trends. Microbiol.* 1:192-196
- Sanger, F., Nicklen, S. and Coulson, A.R.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74(12):5463-5467

- Sathyamoorthy, V. and DasGupta, B.R.** 1985. Partial amino acid sequences of the heavy and light chains of botulinum neurotoxin type E. *Biochem. Biophys. Res. Comm.* **127**(3):768-772.
- Sauer, U., Santangelo, J.D., Treuner, A., Buchholz, M. and Durre, P.** 1995. Sigma factor and sporulation in *Clostridium*. *FEMS Microbiol. Rev.* **17**:331-340.
- Sauer, U., Treuner, A., Buchholz, M., Santangelo, J.D. and Durre, P.** 1994. Sporulation and primary sigma factor homologous genes in *Clostridium acetobutylicum*. *J. Bacteriol.* **176**(21):6572-6582.
- Scherzinger, E., Kruft, V. and Otto, S.** 1993. Purification of the large mobilisation protein of plasmid RSF1010 and characterisation of its site-specific DNA-cleaving/DNA-joining activity. *Eur.J.Biochem.* **217**:929-938
- Scherzinger, E., Lurz, R., Otto, S. and Dobrinski, B.** 1992. *In vitro* cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilisation proteins. *Nucl. Acid. Res.* **20**:41-48
- Schiavo, G., Rossetto, O., Santucci, A., DasGupta, B.R. and Montecucco, C.** 1992. Botulinum neurotoxins are zinc proteins. *J. Biol. Chem.* **267**(33):23479-23483.
- Schiavo, G., Santucci, A., DasGupta, B.R., Mehta, P.P., Jontes, J., Benfenati, F., Wilson, M.C. and Montecucco, C.** 1993. Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct COOH-terminal peptide bonds. *FEBS Lett.* **335**(1):99-103.
- Schmidt, J.J., Sathyamoorthy, V. and DasGupta, B.R.** 1984. Partial amino acid sequence of the heavy and light chains of botulinum neurotoxin type A. *Biochem. Biophys. Res. Comm.* **119**(3):900-904.

- Schmidt, J.J., Sathyamoorthy, V. and DasGupta, B.R.** 1985. Partial amino acid sequences of botulinum neurotoxins types B and E. *Archives Biochem. Biophys.* **238**(2):544-548.
- Scott, A.B.** 1980. Botulinum injection into extraocular muscles as an alternative to strabismus surgery. *Ophthalmology* **87**:1044-1049.
- Scott, V.N. and Duncan, C.L.** 1978. Cryptic plasmids in *Clostridium botulinum* and *C. botulinum*-like organisms. *FEMS Microbiol. Letts.* **4**:55-58.
- Seery, L. T., Nolan, N. C., Sharp, P. M. and Devine, K. M.** 1993. Comparative analysis of the pC194 group of rolling circle plasmids. *Plasmid.* **30**:185-196.
- Sellin, L.C., Thesleff, S. and DasGupta, B. R.** 1983. Different effects of types A and B botulinum toxin on transmitter release at the rat neuromuscular junction. *Acta. Physiol. Scand.* **119**:127-133.
- Shapiro, R.E., Specht, C.D., Collins, B.E., Woods, A.S., Cotter, R.J. and Schnaar, R.L.** 1997. Identification of a ganglioside recognition domain of tetanus toxin using a novel ganglioside photoaffinity ligand. *J. Biol. Chem.* **272**(48):30380-30386.
- Shimizu, T. and Kondo, H.** 1973. Immunological difference between the toxin of a proteolytic strain and that of a nonproteolytic strain of *Clostridium botulinum* type B. *Japan J. Med. Sci. Biol.* **26**:269-271.
- Shimizu, T., Okabe, A. and Rood, J.I.** 1997. Regulation of toxin production in *Clostridium perfringens*. In *The Clostridia: Molecular Biology and Pathogenesis*. P.451-470. Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.

- Shine, J. and Dalgarno, L.** 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to non-sense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA.* **71**:1342-1346.
- Shone, C.C., Hambleton, P. and Melling, J.** 1985. Inactivation of *Clostridium botulinum* type A neurotoxin by trypsin and purification of two tryptic fragments. *Eur. J. Biochem.* **151**:75-82.
- Shone, C.C., Hambleton, P. and Melling, J.** 1987. A 50-kDa fragment from the NH₂-terminus of the heavy subunit of *Clostridium botulinum* type A neurotoxin forms channels in lipid vesicles. *Eur. J. Biochem.* **167**:175-180.
- Siegel, L.S. and Metzger, J.F.** 1979. Toxin production by *Clostridium botulinum* type A under various fermentation conditions. *Appl. Environ. Microbiol.* **38**(4):606-611.
- Siegel, L.S. and Metzger, J.F.** 1980. Effect of fermentation conditions on toxin production by *Clostridium botulinum* type B. *Appl. Environ. Microbiol.* **40**(6):1023-1026.
- Skurnik, M. and Toivanen, P.** 1992. LcrF is the temperature-regulated activator of the *yadA* gene of *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. *J. Bacteriol.* **174**(6):2047-2051.
- Sohling, B. and Gottschalk, G.** 1996. Molecular analysis of the anaerobic succinate degradation pathway in *Clostridium kluyveri*. *J. Bacteriol.* **178**(3):871-880.
- Sollner, T., Whiteheart, S. W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P. and Rothman, J. E.** 1993. SNAP receptors implicated in vesicle targeting and fusion. *Nature* **362**: 318-324

- Soltis, M.T., Mekalanos, J.J. and Betley, M.J.** 1990. Identification of a bacteriophage containing a silent *Staphylococcal* variant enterotoxin gene (*sezA*⁺). *Infect. Immun.* **58**(6):1614-1619.
- Somers, E. and DasGupta, B.R.** 1991. *Clostridium botulinum* types A, B, C₁ and E produce proteins with or without haemagglutinating activity: do they share common amino acid sequences and genes. *J. Prot. Chem.* **10**(4):415-424.
- Southern, E.M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **95**:508-517.
- Soutschek-Bauer, E., Hartl, L. and Staudenbauer, W.L.** 1985. Transformation of *Clostridium thermohydrosulfuricum* DSM 568 with plasmid DNA. *Biotech. Lett.* **7**(10):705-710.
- Spangler, B.D.** 1992. Structure and function of cholera toxin and the related *Escherichia coli* heat-labile enterotoxin. *Microbiol. Rev.* **56**(4):622-647
- Sprengel, R., Reiss, B. and Schaller, H.** 1985. Translationally coupled initiation of protein synthesis in *Bacillus subtilis*. *Nucl. Acids. Res.* **13**(3):893-909.
- Squires, C.H., Heefner, D.L., Evans, R.J., Kopp, B.J. and Yarus, M.J.** 1984. Shuttle plasmids for *Escherichia coli* and *Clostridium perfringens*. *J. Bacteriol.* **159**(2):465-471.
- Stecher, B., Gratzl, M. and Ahnert-Hilger, G.** 1989. Reductive chain separation of botulinum A toxin - a prerequisite to its inhibitory action on exocytosis in chromaffin cells. *FEBS Lett.* **248**(1,2):23-27.
- Stevens, D. L.** 1997. Necrotizing clostridial soft tissue infections. In: *The Clostridia: Molecular Biology and Pathogenesis*. P.141-151. Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.

- Stewart, G.J. and Carlson, C.A.** 1986. The biology of natural transformation. *Ann. Rev. Microbiol.* **40**:211-235.
- Strom, M.S., Eklund, M.W. and Poysky, F.T.** 1984. Plasmids in *Clostridium botulinum* and related *Clostridium* species. *Appl. Environ. Microbiol.* **48**(5):956-963.
- Sugii, S., Ohishi, I. And Sakaguchi, G.** 1977. Correlation between oral toxicity and *in vitro* stability of *Clostridium botulinum* type A and type B toxins of different molecular sizes. *Infect. Immun.* **16**:910-914.
- Sugii, S., Ohishi, I. And Sakaguchi, G.** 1977. Intestinal absorption of botulinum toxins of different molecular sizes in rats. *Infect. Immun.* **17**:491-496
- Sugiyama, H.** 1980. *Clostridium botulinum* neurotoxin. *Microbiol. Rev.* **44**(3):419-448.
- Sunagawa, H. and Inoue, K.** 1991. Isolation and characterisation of converting and non-converting phages, harbored in the strains of *Clostridium botulinum* types C and D isolated in Japan. *J. Vet. Med. Sci.* **53**(5):951-954.
- Swinfield T-J.** 1987. Development of host:vector systems in *Clostridium*. Ph.D Thesis. University of Newcastle-Upon-Tyne.
- Takumi, K. and Kawata, T.** 1970. Chemical composition of the cell walls of *Clostridium botulinum* type A. *Japan J. Microbiol.* **14**(1):57-63.
- Takumi, K., Kawata, T. and Hisatsune, K.** 1971. Autolytic enzyme system of *Clostridium botulinum*. II. Mode of action of autolytic enzymes in *Clostridium botulinum* type A. *Japan J. Microbiol.* **15**(2):131-141.
- Tesh, V.L. and O'Brien, A.D.** 1991. The pathogenic mechanisms of Shiga toxin and the Shiga-like toxins. **5**(8):1817-1822.

- Thanabalu, T., Hindley, J. and Berry, C.** 1992. Proteolytic processing of the mosquitocidal toxin from *Bacillus sphaericus* SSII-1. *J. Bacteriol.* **174**(15):5051-5056.
- Thomas, Jr.W.D. and Archer, G.L.** 1992. Mobilisation of recombinant plasmids from *Staphylococcus aureus* into coagulase negative *Staphylococcus* species. *Plasmid.* **27**:164-168.
- Thompson, D.E., Brehm, J.K., Oultram, J.D., Swinfield, T-J., Shone, C.C., Atkinson, T., Melling, J. and Minton, N.P.** 1990. The complete amino acid sequence of the *Clostridium botulinum* type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene. *Eur. J. Biochem.* **189**:73-81.
- Tindal, K. R. and Kunkel. T. A.** 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry* **27**:6008-6013.
- Titball, R. W.** (1997) Clostridial phospholipases. In *The Clostridia: Molecular Biology and Pathogenesis*. P.223-242 Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Titball, R.W.** 1993. Bacterial phospholipases C. *Microbiol. Rev.* **57**(2):347-366.
- Tomizawa, J-I.** 1990. Control of ColE1 plasmid replication: Interaction of Rom protein with an unstable complex formed by RNA I and RNA II. *J. Mol. Biol.* **212**:695-708.
- Trevors, J.T.** 1991. Electrotransformation of bacteria. *Methods in Molecular Cellular Biology* **2**:247-253.
- Truffaut, N. and Sebald, M.** 1983. Plasmid detection and isolation in strains of *Clostridium acetobutylicum* and related species. *Mol. Gen. Genet.* **189**:178-180.
- Truffaut, N., Hubert, J. and Reysset, G.** 1989. Construction of shuttle vectors useful for transforming *Clostridium acetobutylicum*. *FEMS Microbiol. Lett.* **58**:15-20.

- Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N. and Oguma, K. 1992.** The complete nucleotide sequence of the gene coding for the nontoxic-nonhemagglutinin component of *Clostridium botulinum* type C progenitor toxin. *Biochem. Biophys. Res. Comm.* **183(2):1273-1279.**
- Tsuzuki, K., Kimura, K., Fujii, N., Yokosawa, N., Indoh, T., Murakami, T. and Oguma, K. 1990.** Cloning and complete nucleotide sequence of the gene for the main component of haemagglutinin produced by *Clostridium botulinum* type C. *Infect. Immun.* **58:3173-3177.**
- Tsuzuki, K., Yokosawa, N., Syuto, B., Ohishi, I., Fujii, N., Kimura, K. and Oguma, K. 1988.** Establishment of a monoclonal antibody recognising an antigenic site common to *Clostridium botulinum* type B, C1, D and E toxins and tetanus toxin. *Infect. Immun.* **56(4):898-902.**
- Tweten, R. K. 1997.** The thiol-activated clostridial toxins. In *The Clostridia: Molecular Biology and Pathogenesis*. P.211-221. Eds. Rood, J. I., McClane, B. A., Songer, J. G. and Titball, R. W. Academic Press Inc. U.S.A.
- Urano, N., Karube, I., Suzuki, S., Yamada, T., Hirochika, H and Sakaguchi, K. 1983.** Isolation and partial characterisation of large plasmids in hydrogen-evolving bacterium *Clostridium butyricum*. *Eur. J. Microbiol. Biotechnol.* **17:349-354.**
- Van der Lelie, D., Bron, S., Venema, G. and Oskam, L. 1989.** Similarity of minus origins of replication and flanking open reading frames of plasmids pUB110, pTB913 and pIMV158. *Nucl. Acids. Res.* **17(18):7283-7295.**

- Vellanoweth, R.L.** 1993. Translation and its regulation. In *Bacillus subtilis* and other Gram-positive organisms. p.699-711 Eds. Sonenshein, A.L., Hoch, J.A. and Losick, R.. A.S.M. Washington D.C.
- Wagner, E.G.H. and Simons, R.W.** 1994. Antisense RNA control in bacteria, phages and plasmids. *Ann. Rev. Microbiol.* **48**:713-742.
- Wagner, L. A., Gesteland, R. F., Dayhuff. T. J. and Weiss, R. B.** 1994. An efficient Shine-Dalgarno sequence but not translation is necessary for *lacZ* mRNA stability in *Escherichia coli*. *J. Bacteriol.* **176**(6):1683-1688.
- Wang, A. and Macrina, F.L.** 1995. Characterisation of six linked open reading frames necessary for pIP501-mediated conjugation. *Plasmid.* **34**:206-210.
- Waters, V.L. and Guiney, D.G.** 1993. Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. *Mol. Microbiol.* **9**(6):1123-1130.
- Weaver, J.C. and Chizadzhev, Y.A.** 1996. Theory of electroporation: A review. *Bioelectrochem. Bioenerget.* **41**:135-160
- Weaver, J.C. and Chizmadzhev, Y.A.** 1996. Theory of electroporation: a review. *Bioelectrochem. Bioenerget.* **41**:135-160.
- Weber, J. T., Goodpasture, H. C., Alexander, H., Werner, S. B., Hatheway, C. L. and Tauxe, R.V.** 1993. Wound botulism in a patient with tooth abcess: case report and review. *Clin. Infect. Dis.* **16**:635-639
- Weickert, M.J., Chambliss, G.H. and Sugiyama, H.** 1986. Production of toxin by *Clostridium botulinum* type A strains cured of plasmids. *Appl. Environ. Microbiol.* **51**(1):52-56.

- Whelan, S. M. (1993)** Molecular characterisation and detection of clostridial neurotoxin genes. Ph.D. Thesis. Open University
- Whelan, S.M., Elmore, M.J., Bodsworth, N.J., Brehm, J.K., Atkinson, T. and Minton, N.P. 1992.** Molecular cloning of the *Clostridium botulinum* structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence. Appl. Environ. Microbiol. **58**(8):2345-2354.
- Whitmer, M.E. and Johnson, E.A. 1988.** Development of improved defined media for *Clostridium botulinum* serotypes A, B and E. Appl. Environ. Microbiol. **54**(3):753-759.
- Wilkinson, S.R., Young, D.I., Morris, J.G. and Young, M. 1995.** Molecular genetics and the initiation of solventogenesis in *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052. FEMS Microbiol. Rev. **17**:275-285.
- Williams, D.R., Young, D.I. and Young, M. 1990.** Conjugative plasmid transfer from *Escherichia coli* to *Clostridium acetobutylicum*. J. Gen. Microbiol. **136**:819-826.
- Winans, S.C., Burns, D.L. and Christie, P.J. 1996.** Adaptation of a conjugal transfer system for the export of pathogenic macromolecules. Trends in Microbiol. **4**(2):64-68
- Woodcock, D. M., Crowther, P. J., Doherty, J., Jefferson, S., DeCruz, E., Noyer-Weidner, M., Smith, S.S., Michael, M. Z. and Graham, M. W. 1989.** Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants. Nucl. Acids. Res. **17**(9):3469-3478.
- Woods, D R. and Reid, S.J. 1995.** Regulation of nitrogen metabolism, starch utilisation and the β -hbd-adh1 gene cluster in *Clostridium acetobutylicum*. FEMS Microbiol. Rev. **17**:299-306.

- Woolley, R.C., Pennock, A., Ashton, R.J., Davies, A. and Young, M.** 1989. Transfer of Tn1545 and Tn916 to *Clostridium acetobutylicum*. Plasmid 22:169-174.
- Wright, J.F., Pernollet, M., Reboul, A., Aude, C. and Colomb, M.G.** 1992. Identification and partial characterisation of a low affinity metal binding site in the light chain of tetanus toxin. J. Biol. Chem. 267(13):9053-9058.
- Yang, G-H., Rhee, S-D., Jung, H-H. and Yang, K-H.** 1996. Organisation and nucleotide sequence of genes for hemagglutinin components of *Clostridium botulinum* type B progenitor toxin. Biochem. Mol. Biol. Int. 39(6):1141-1146.
- Yanisch-Perron, C., Vieria, J. and Messing, J.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene. 33:103-119.
- Yoshino, S., Yoshino, T., Hara, S., Ogata, S. and Hayashida, S.** 1990. Construction of a shuttle vector plasmid between *Clostridium acetobutylicum* and *Escherichia coli*. Agric. Biol. Chem. 54(2):437-441.
- Young, M.** 1993. Development and exploitation of conjugative gene transfer in Clostridia. In The Clostridia and Biotechnology. P.99-117 Ed. Woods, D.R. Butterworth-Heinemann. U.S.A.
- Young, M., Minton, N.P. and Staudenbauer, W.L.** 1989. Recent advances in the genetics of the clostridia. FEMS. Microbiol. Rev. 5(4):301-325.
- Yudkin, M.D.** 1987. Structure and function in a *Bacillus subtilis* sporulation-specific sigma factor: Molecular nature of mutations in *spoIIAC*. J. Gen. Microbiol. 133:475-481
- Zhou, Y. and Johnson, E.A.** 1993. Genetic transformation of *Clostridium botulinum* Hall A by electroporation. Biotech. Letts. 15(2):121-126.

Zhou, Y., Sugiyama, H. and Johnson, E.A. 1993. Transfer of neurotoxigenicity from *Clostridium butyricum* to a nontoxigenic *Clostridium botulinum* type E-like strain. Appl. Environ. Microbiol. **59**(11):3825-3831.

Zhou, Y., Sugiyama, H., Nakano, H. and Johnson, E.A. 1995. The genes for the *Clostridium botulinum* type G toxin complex are on a plasmid. Infect. Immun. **63**(5):2087-2091.

Zuker, M. 1989. On finding all suboptimal foldings of an RNA molecule. Science. **244**:48-52.

APPENDIX I

BUFFERS AND STOCK SOLUTIONS

Where appropriate solutions were diluted to working concentration with sterile deionised water. Autoclaving was at 121 °C, 15 lb/sq. in. for 20 min.

Acrylamide stock (40 %)

Acrylamide	380	g
Bis-acrylamide	20	g
Made up to 1 litre with deionised H ₂ O and deionised with amberlite MB-1 resin and filtered.		

Acrylamide gel mix A (0.5xTBE)

40 % acrylamide stock	150	ml
10x TBE buffer	50	ml
Urea (Ultra pure)	460	g
Made up to 1 litre with deionised H ₂ O.		

Acrylamide gel mix B (5.0x TBE)

40 % acrylamide stock	150	ml
10x TBE buffer	500	ml
Urea (Ultra pure)	460	g
Bromophenol blue	0.05	g
Made up to 1 litre with deionised H ₂ O.		

Brij/DOC (pH 8.0)

Tris-HCl	50	mM
Brij-58	1.0 %	(w/v)
Na deoxycholate	0.4 %	(w/v)
EDTA	25	mM

Chase solution

dATP	2.5	mM
dTTP	2.5	mM
dGTP	2.5	mM
dCTP	2.5	mM

Made from 50 mM stock solutions with 1xTE.

AE buffer(1x)(pH8.0)

Tris-HCl	50	mM
EDTA	10	mM

PEB buffer (3x)

PIPES	30	mM
EDTA	1.5	mM
NaCl	1.2	M

DNase I buffer

Sodium acetate	100 mM
MgSO ₄	5 mM
pH 5.0	

DNA tracking dye

Bromophenol blue	0.25 g/100 ml
Sucrose	40 g/100 ml

Formaldehyde gel-loading buffer (pH 8.0)

Glycerol	50% (v/v)
EDTA	1 mM
Bromophenol blue	0.25 % (w/v)
Xylene cyanol FF	0.25 % (w/v)

GOOP prehybridisation buffer (pH 6.5)

SDS	70.0 g/L
EDTA	1.86 g/L
NaH ₂ PO ₄	41.0 g/L
Na ₂ HPO ₄	41.0 g/L

GTEL buffer (pH8.0)

Glucose	50 mM
Tris	25 mM
EDTA	10 mM

Lysis solution

SDS	1 % (w/v)
NaOH	0.2 M

Ligation buffer (10x) (pH 7.5)

Tris-HCl	300 mM
NaCl	300 mM
MgCl ₂	75 mM
Spermidine	10 mM
ATP	2.5 mM
DTT	20 mM
Na ₂ EDTA	20 mM

Kinase buffer(10x) (pH 7.4)

Tris-HCl	500 mM
MgCl ₂	100 mM
DTT	50 mM

S.E.B electroporation buffer (pH 7.4)		
Sucrose	270	mM
Na ₂ HPO ₄ /NaH ₂ PO ₄	7	mM
MgCl ₂	1	mM
SSC buffer (20x) (pH 7.0)		
NaCl	3.0	M
Na ₃ citrate.2H ₂ O	0.3	M
Strip neutralisation buffer (pH 7.5)		
NaCl	1.5	mM
Tris-HCl	0.5	M
S.T.E.T buffer (pH 8.0)		
Sucrose	8 g/100 ml	
Triton X-100	5 ml/100 ml	
Na ₂ EDTA	0.005	M
Tris-Hcl	0.05	M
TE buffer (1x) (pH 8.0)		
Tris HCl	10	mM
Na ₂ EDTA	0.1	mM
TBE buffer (10x) (pH 8.3)		
Tris-borate	0.9	M
Boric acid	0.9	M
Na ₂ EDTA	0.03	M
TM buffer (1x) (pH 8.5)		
Tris-HCl	100	mM
MgCl ₂	50	mM
AMV reverse transcriptase buffer (x5) (pH 8.5)		
Tris-HCl	250	mM
MgCl ₂	40	mM
KCl	150	mM
Dithiothreitol	5	mM
Calf intestinal alkaline phosphatase buffer (x10)(pH 8.5)		
Tris-HCl	0.5	M
EDTA	1.0	mM

Amplitaq DNA polymerase buffer (x10)(pH 8.3)

MgCl ₂	15 mM
KCl	500 mM
Tris-HCl	100 mM
Gelatin	0.1 %

T4 polynucleotide kinase buffer (x5)(pH 7.6)

Tris-HCl	350 mM
KCl	500 mM
2-Mercaptoethanol	5 mM
MgCl ₂	50 mM

Tris-Sucrose solution (pH 8.0)

Tris-HCl	50 mM
Sucrose	25 % w/v

DNA sequencing reaction mixes**Labeling mix**

dCTP	37.5 μM
dGTP	37.5 μM
dTTP	37.5 μM

Stock solutions diluted in sterile deionised H₂O.

Labeling mixture for 1 reaction

1x TE	2.0 μL
DTT(100 mM)	1.0 μL
H ₂ O	1.6 μL
T7 polymerase	0.5 μL
[α- ³⁵ S]dATP	0.5 μL

Dideoxy termination mixtures

'A' ddATP(4 μM/L), dWTPs(160 μM/L), dSTPs(80 μM/L), 50 mM NaCL

'T' ddTTP(4 μM/L), dWTPs(160 μM/L), dSTPs(80 μM/L), 50 mM NaCL

'C' ddCTP(8 μM/L), dWTPs(160 μM/L), dSTPs(80 μM/L), 50 mM NaCL

'G' ddGTP(8 μM/L), dWTPs(160 μM/L), dSTPs(80 μM/L), 50 mM NaCL

Antibiotic supplements**Ampicillin stock solution**

100 mg of ampicillin sodium salt dissolved per 1 ml of deionised H₂O and sterilised by filtration using a Sartorius 0.22 μm disposable filter.

Working concentration: 50-100 μg/ml.

Storage -20 °C.

Erythromycin stock solution (*E. coli*)

100 mg of erythromycin dissolved per 1 ml of 70 % ethanol and sterilised by filtration using a Sartorius 0.22 µm disposable filter.

Working concentration: 500 µg/ml.

Storage -20 °C.

Erythromycin stock solution (*Clostridium*)

10 mg of erythromycin dissolved per 1 ml of 70 % ethanol and sterilised by filtration using a Sartorius 0.22 µm disposable filter.

Working concentration: 5-50 µg/ml.

Storage -20 °C.

Kanamycin stock solution

10 mg of kanamycin sulphate dissolved per 1 ml of deionised H₂O and sterilised by filtration using a Sartorius 0.22 µm disposable filter.

Working concentration: 50 µg/ml.

Storage -20 °C.

APPENDIX II

MEDIA

***E.coli* MEDIA**

Luria-Bertani medium (pH 7.4)

	g/l
Tryptone	10.0
Yeast extract	5.0
NaCl	5.0
Agar (solid media)	20.0

2x YT (pH 7.4)

	g/l
Tryptone	16.0
Yeast extract	10.0
NaCl	5.0
Agar (solid media)	20.0

SOC medium (pH 7.5)

	g/l	
Tryptone	20.0	
Yeast extract	5.0	
NaCl	20.0	mM
KCl	2.5	mM
MgCl ₂	10.0	mM
MgSO ₄	10.0	mM
Glucose	20.0	mM

H-Top Agar

	g/l
NaCl	8.0
Agar	8.0

M9 media (x10)(pH 7.4)

M9 Salts	g/l
Na ₂ HPO ₄ .7H ₂ O	128.0
KH ₂ PO ₄	30.0
NaCl	5.0
NH ₄ Cl	10.0

Non-autoclaved requirements.

	ml/l
1 M MgSO ₄	2.0
0.1 M CaCl ₂	0.2
20 % glucose	5.0
Amino-acids	2.0

M9 agar plates were made by initially autoclaving a 2 % (w/v) agar solution. This was cooled to @ 50 °C and 100 ml of 10x M9 salts added. Sterile solutions of MgSO₄, CaCl₂ ,

glucose and specific requirements, typically amino acids were then added using an aseptic technique.

CLOSTRIDIAL MEDIA

USA II (pH 7.3)

	g/l
Proteose peptone	20.0
Yeast extract	10.0
N.Z. Amine A	10.0
Sodium thioglycollate	0.5
Agar (solid media)	20.0

YTG Media (pH 7.4)

	g/l
Tryptone	16.0
Yeast extract	10.0
NaCl	5.0
Glucose	5.0
Agar (solid media)	20.0

TPGY media

	g/l
Tryptone	20.0
Peptone	5.0
Glucose	1.0
Yeast Extract	5.0
L-Cysteine HCl	1.0
Agar (solid media)	20.0

CMC broth

Trypicase peptone	25	g
Yeast extract	5.0	g
Neutralised liver digest	5.0	g
L-Cysteine HCl	0.5	g
Sodium formaldehyde-sulphoxylate	0.3	g
K ₂ HPO ₄ (anhydrous)	5.0	g
Vitamin K1 solution	1.0	ml
Haemin solution	10	ml
Glucose	4.0	g
Soluble starch	1.0	g
Cellobiose	1.0	g
Maltose	1.0	g
Resazurin	1.0	mg/ml
Millipore H ₂ O	1000	ml

APPENDIX III

**Annotated nucleotide sequence of neurotoxin complex operons
5' to the silent type B neurotoxin in *C. botulinum* NCTC 2916**

Hind III

AAGCTTACATTATCAGATAAGTATTTTTTACAATTTCTAATTTTAATTCAGTTGAATATATTTTTTTTGCCACAAAAATGCACCTCCAA

TTCGAATGTAATAGTCTATTTCATAAAAAATGTTAAAGATTAAAATTAAGTCAACTTATATAAAAAAACGGTGTTTTACGTGGAGGTT

90

AAATTTAGATTTTTGTCTAAACTTTGGGGTGCACCTTCATAAGCTGGTGGACTCTTATTTAATAATAAATCGTAATATATTAATTAGTAAT

TTTAAATCTAAAAACAGATTGAAACCCACGTGAAGTATTCGACCACCTGAGAATAAATTATTATTTAGCATTATATAATTAATCATT
N T I

180

ATCTATATGCAATCTTATATTATAGTTATTTAATTCTGTAACCTCTACTTTTAATATATAAATTGCACCATTTAATAAACGTATTGAATT

TAGATATACGTTAGAATATAATATCAATAAATTAAGACATTGAAGATGAAAATTATATATTTAACGTGGTAAATTATTTGCATAACTTAA
D I H L R I N Y N N L E T V E V K L I Y I A G N L L R I S N

270

GGTATTATTTAAGAGGTGTAAATTTTGTCTATACCAATTAACCTTATAGCCGTCAATAGTTTCAATAACTCTAAATAAAGTACCTATTCC

CCATAATAAATTCTCCACATTTAAACAGCATATGGTTAATTGAATATCGGCAGTTATCAAAGTTATTGAGATTTATTTTCATGGATAAGG
T N N L L H L N Q R I G I L K Y G D I T E I V R E L T G I G

360

TTGATTATTACTAGTATAAAGCCTAATGTTAGATGTATCTTTATTATTAAAGTTACCAGGCACTGTAAATTCATATGAAATATAGTTACC

AACTAATAATGATCATATTTCCGATTACAATCTACATAGAAATAATAATTTCAATGGTCCGTGACATTTAAGTATACTTTATATCAATGG
Q N N S T Y L R I N S T D K N N F N G P V T F E Y S I Y N G

450

AACATCAGGGCTTTGTCTTGTATAGTAATGTGCACCTGTTGAATTAAGATTATCTATTATATTAACCTGAAGAATTCATTAAATTTCCGGAG

TTGTAGTCCCGAAACAGAACATATCATTACACGTGGACAACTTAATTCTAATAGATAATATAATTGACTTCTTAAGTAATTTAAAGCCTC
V D P S Q R T Y Y H A G T S N L N D I I N V S S N M L N R L

540

ATCCCCTTCTTCTCTCTTGAACCTTTATTTGATTTTCTATATTAGGAGAAGTTTGTGGTACTCTAATGTAGTAATTTTTATCCTTATTAAG

TAGGGGAAGAAGAGAGAACTTGAAATAAACTAAAAGATATAATCCTCTTCAAACACCATGAGATTACATCATTAAAAATAGGAATAATTC
D G E E R K F K I Q N E I N P S T Q P V R I Y Y N K D K N L

630

CAAAACAATAAATAACTTTTAGTATTAGGGTCGTCAACCCAGTAATATAATTTATTGCTTTAATATTATTATTACATAAATTGCCAGC

GTTTTGTTGATTTATTGAAAATCATAATCCCAGCAGTTGGGGTCATTATATTAATAACGAAATTATAATAATAATGTATTTAACGGTCG
L V V L Y S K T N P D D V G T I Y N I A K I N N N C L N G A

720

TTGCAAATTTCTATTACCCGAACCTATCGCTTCATATATTTTAAATGGTGGTAGATTATTTTGAGTATTAATTTTATAAATTCCTGTAGT

AACGTTTAAAGATAATGGGCTTGGATAGCGAAGTATATAAAAATTACCACCATCTAATAAACTCATAATTAAATATTTAAGGACATCA
Q L N R N G S G I A E Y I K L P P L N N Q T N I K Y I G T T

810

ATTTACTGTAAATAATATATAAAATGGATTATTATTTACAATACCTGGAATGTTATATGTGTAATTATCTGAAATAGCAGTATTTATTTT

TAAATGACATTTATTATATATTTTACCTAATAATAAATGTTATGGACCTTACAATATACACATTAATAGACTTTATCGTCATAAATAAAG
N V T F L I Y F P N N N V I G P I N Y T Y N D S I A T N I E

900

TTGTATATTTTCAGATGGGTCTATGACACCTACTACAATTTCTTCCTCGGGACCATTTTGTATAAATCCATCTTCAGCCACTTGCATTTT
 ----- 990
 AACATATAAAAGTCTACCCAGATACTGTGGATGATGTTAAAGAAGGAGCCCTGGTAAAACATATTTAGGTAGAAGTCGGTGAACGTAAAA
 Q I N E S P D I V G V V I E E E F G N Q I F G D E A V Q M K

 AACTACTTCAGATTCAGGTGAAGGAGCATTAGTTACTTCACAATATCCTAAGGAAGGAAGATAATACTCAGCTTTATTAACAATACTTCC
 ----- 1080
 TTGATGAAGTCTAAGTCCACTTCCTCGTAATCAATGAAGTGTATAGGATTCCTTCCTTCTATTATGAGTCGAAATAATTGTTATGAAGG
 V V E S E F S F A N T V E C Y G L S F L Y Y E A K N V I S G

 ATTTTGATAGTCTATTAATTCATATTTTTCATAGGTATTATAAGCCTTTATGCATATATTTTATCTGAAGGAGCTTCTTCTACAAATTT
 ----- 1170
 TAAACTATCAGATAATTAAGTATAAAAAGTATCCATAATATTCGGAATACGTATATAAAAATAGACTTCCTCGAAGAAGATGTTTAAA
 N Q Y D I L E Y K E Y T N Y A K I C I N K D S F A E E V F K

 AAATAACGATGTTGATGTATTAAATAATTGTACATTGAAAGCGTCTTTAAGTCTTTCCTCAGTATATTGTTGTGAGAATTTTGTATCATT
 ----- 1260
 TTTATTGCTACAACCTACATAATTTATTAACATGTAACCTTTCGAGAAATTCAGAAAGAAGTCATATAACAACACTCTTAAACATAGTAA
 F L S T S T N F L Q V N F A D K L R E E T Y Q Q S F K T D N

 ACTTGTGGTGAATAATGGTCTAGTTGTATTTCTTAATCGAGGTTGTATAAATCTTCCTGATGAACCTGCTTCGATTAATAATGTACCTAT
 ----- 1350
 TGAACACCACTTATTACCAGATCAACATAAAGAATTAGCTCCAACATATTATGAAGGACTACTTGGACGAAGCTAATTATTACATGGATA
 S T T F L P R T T N R L R F Q I I S G S S G A E I L L T G I

 CAAATCTTTATCATTGTTCTTATATATCCATCACCTTATTAATAACATATAGTCCATTACTATAAGGCAATACTCGTTGTGTGCTACT
 ----- 1440
 GTTTAGAAATAGTAAACAAGAATATATAGGTAGTGGGAATAATTATTGTATATCAGGTAATGATATTCCGTTATGAGCAACACACGATGA
 L D K D N T R I Y G D G K N I V Y L G N S Y F L V R Q T S S

 AACTTTAGAATTAAAAATATTGTTAGTTGGTATATTTTAGTAGTTTCTAAAATTTTTTATATAGGAATTGTTCAATTTAAATGCACTT
 ----- 1530
 TTGAAATCTTAATTTTTTATAACATCAACATATAAAATCATCAAAGATTTTAAAAAATATATCCTTAACAAGTAAATTTTACGTGAA
 V K S N F I N N T P I N K T T E L I K K Y L F Q E N L I C K

 ATCATTAGTTAGAAAGTAAAAATCTTTATTTATTTTTTCCCCTGGATTAAATTCTATAGACTTAACATATCCAAGAGATGGAACATATAA
 ----- 1620
 TAGTAATCAATCTTTCATTTTTAGAAATAAATAAAAAAGGGGACCTAATTTAAGATATCTGAATTGTATAGGTTCTCTACCTTGTATATT
 D N T L F Y F D K N I K E G P N F E I S K V Y G L S F V Y L

 AACTGCTCTCTCTATAATTTCATGTTGTAAGACTTTTATTATTTTCATATCTAATGTAGGTATATTGTAAATACATATATAAGTTTTTTATT
 ----- 1710
 TTGACGAGAGAGATATTAAAGTACAACATTCTGAAAATAATAAAGTATAGATTACATCCATATAACATTTATGTATATATTCAAAAATAA
 V A R E I I E H Q L V K I I E Y R I Y T Y Q L Y M Y L N K N

 TGAGGGAGCGGTTTTACTAAATTCAAATCCTATTGGAATTTGATTAGCTTCAGTAAAGTTAGTAAATACAGTTTGTATATTATTTTTTAT
 ----- 1800
 ACTCCCTCGCCAAAATGATTTAAGTTTAGGATAACCTTAACTAATCGAAGTCATTTCAATCATTTATGTCAAACATATAATAAAAAATA
 S P A T K S F E F G I P I Q N A E T F N T F V T Q I N N K I

ATATTCTTCATTGAAAGATGGTGTGGATAATAATATGGTATCGCATTATCATTTACACGTAGGTCCCCAACTATTCCTGTTGATCCATT
 ----- 1890
 TATAAGAAGTAACTTTCTACCACAACCTATTATTATACCATAGCGTAATAGTAAATGTGCATCCAGGGGTTGATAAGGACAACCTAGGTAA
 Y E E N F S P T P Y Y Y P I A N D N V R L D G V I G T S G N
XbaI
 ACTAATTACACTTCCACCTAATATTTGATTTTGTCTAGATAATATCCATCCATCCCCTCTTCTAACTACATAATTACCATCAGCTAAATC
 ----- 1980
 TGATTAATGTGAAGGTGGATTATAAACTAAAACAGATCTATTATAGGTAGGTAGGGGAGAAGATTGATGTATTAATGGTAGTTCGATTTAG
 S I V S G G L I Q N Q R S L I W G D G R R V V Y N G D A L D
EcoRI
 AATAGTATCACTATAGTTTATAACTTTTTCTTGTATATCATTATAAATTTTTTTTATAGATGAATTCATACACTTTTCTCCTTTATATTT
 ----- 2070
 TTATCATAGTGATATCAAATATTGAAAAAGAACATATAGTAATATTTAAAAAAATATCTACTTAAGTATGTGAAAAGAGGAAATATAAA
 I T D S Y N I V K E Q I D N Y I K K I S S N M ← HA70 I K
 TTTCAAGTTTGAACATTTGATTTGAATTGTATATATCAAATTCGGTAATAGGAGTAGTGGTTGGCTTAAAATATTTTCATTAGTATCAT
 ----- 2160
 AAAGTTCAAACCTTGTAACCTAACTTAACATATATAGTTTAAAGCCATTATCCTCATCACCAACCGAATTTTATAAAAGTAATCATAGTA
 E L K F M Q N S N Y I D E N P L L L L P Q S L I N E N T D Y
 AAATGTCCCAGGCATAATCTAATTCATTCACTTTATTTAACTTAACATAATATAAGTATTTACAGCAATTTTAATAGGAAACCAGTAGC
 ----- 2250
 TTTACAGGGTCCGTATTAGATTAAGTAAGTGAAATAAATTTGAATTGTATTATATTCATAAATGTCGTTAAAATTATCCTTTGGTCATCG
 I D W A Y D L E N V K N L S L M I Y T N V A I K I P F W Y C
 ATCTATTGGATAATGAATCTAAAGAAATAAATCCAAAGTTATCGTAACTTAAATACTTATTTGGTTCTGCTACATTAGAGATTTTAAAGC
 ----- 2340
 TAGATAACCTATTACTTAGATTTCTTTATTTAGGTTTCAATAGCATTGAATTTATGAATAAACCAAGACGATGAATCTCTAAATTTTCG
 R N S L S D L S I F G F N D Y S L Y K N P E A V N S I K F C
TaqI
 ATCTATTTTCAGCCATATATTCTACATTCCATTTTGGATTATTTGCAGAAGATTCATTTGAAAATGTTAATGATTTGATACAGGATTTA
 ----- 2430
 TAGATAAAAGTCGGTATATAAGATGTAAGGTAAAACTAATAAACGTCTTCTAAGTAACTTTTACAATTACTAAAGCTATGTCCTAAAT
 R N E A M Y E V N W K Q N N A S S E N S F T L S K S V P N L
 AATATAAAGAACCAGAAAAGATAGATTTTATATTGTAATTACCATTAGGTAGAAAAGTTCTTTCAACTGACATATAATTCACCTCTTTAT
 ----- 2520
 TTATATTTCTTGGTCTTTTCTATCTAAAATATAACATTAATGGTAATCCATCTTTTCAAGAAAGTTGACTGTATATTAAGTGGAGAAATA
 Y L S G S F I S K I N Y N G N P L F T R E V S M ← HA17
ScaI
 AAATTAGATGATAGTACTCATATTAAATCTTTGTTGATATTAATTTTATGGGTTACGAATATTCATTTCTGATTATCATCTCCATGATA
 ----- 2610
 TTTAATCTACTATCATGAGTATAATTTAGAAACAACTATAATTTAAATACCCAATGCTTATAAGGTAAAGACTAATAGTAGAGGTACTAT
 P N R I N W K Q N D D G H Y
XbaI
 ATTAAATACTTGAATAGCAGTTCCGTTTGCTGTTTGGCTGTTATATAAATCTAGAGCTTTAGTTGTATCGCGTAGATTAGTAATTGTATA
 ----- 2700
 TAATTTATGAACTTATCGTCAAGGCAAACGACAAACCGACAATATATTTAGATCTCGAAATCAACATAGCGCATCTAATCATTAAACATAT
 N F V Q I A T G N A T Q S N Y L D L A K T T D R L N T I T Y

TGTTTCATCAGTATCTGAAACAGGATTTATAAGCCAATATTGGGCGTCATTATTTTGATCATTAGAGAAGAAACCCTTACAGTATTACC 2790

 ACAAAGTAGTCATAGACTTTGTCCTAAATATTCGGTTATAACCCGCAGTAATAAACTAGTAATCTTCTTCTTTGGGAATGTCATAATGG
 T E D T D S V P N I L W Y Q A D N N Q D N S S S V R V T N G

 ATTTGAAAAAATCCATGTTAGAACTCCGTTTGAAAGTATTGTATTAAAAACTGGTATGCTGCTTTTTCTTCATTATATCTAATTGTCCA 2880

 TAACTTTTTTAGGTACAATCTTGAGGCAAACCTTTCATAACATAATTTTTTGACCATACGACGAAAAAGAAGTAATATAGATTAAACAGGT
 N S F I W T L V G N S L I T N F F Q Y A A K E E N Y R I T W

 TTTTTGATTGCGACCATAGTCCCAAGTATATAAATTAACATTTAGATTTGTCATATCCACTTGTTGTACGACTTTATTACGATCTAATAT 2970

 AAAAATAACGCTGGTATCAGGGTTCATATATTTAATTGTAAATCTAAACAGTATAGGTGAACAACATGCTGAAATAATGCTAGATTATA
 K Q N R G Y D W T Y L N V N L N T M D V Q Q V V K N R D L I

Bgl II
 TGGACTTATTTTACATGTGAAATTGTTAAGATCTGATATTATATAATCTTCTATGATAAATTTTATATAATTTGAATTATTAAGTGTGCT 3060

 ACCTGAATAAAATGTACACTTTAACAATTCTAGACTATAATATATTAGAAGATACTATTTAAAATATATTAACTTAATAATTCACACGA
 P S I K C T F N N L D S I I Y D E I I F K I Y N S N N L T S

Hind III
 AAGCTTCAAATTACGAGCTACGGTATCAGCATATAATACTAAGTTAGGGTTTTTATAACTTGCAATAATAAATGAATTGTTACCAATGTC 3150

 TTCGAAGTTTAATGCTCGATGCCATAGTCGTATATTATGATTCAATCCCAAAAATATTGAACGTTATTATTTACTTAACAATGGTTACAG
 L K L N R A V T D A Y L V L N P N K Y S A I I F S N N G I D

 TTTTAATAATAACCAATATTGATTATCTGCATTGGAATCTTGTTGCGTTGATATATTATGTGTTGGTGCATTCCATGTTAAAACTAAATT 3240

 AAAATTATTATTGGTTATAACTAATAGACGTAACTTAGAACAACGCAACTATATAATACACAACCACGTAAGGTACAATTTTGATTAA
 K L L L W Y Q N D A N S D Q Q T S I N H T P A N W T L V L N

EcoR V
 AGTATTATGGATATCCATACTTTTTATTTTATAAGCAGCTTTATTAGAATCATATATAAGTCTCCATCTTTCAAGGTAATTTCTAGTTG 3330

 TCATAATACCTATAGGTATGAAAAATAAAATATTCGTCGAAATAATCTTAGTATATATTTCAGAGGTAGAAAGTTCCATTAAAGATCAAAC
 T N H I D M S K I K Y A A K N S D Y I L R W R E L Y N R T Q

Hae III
 TTGAAATAAGCTAACGTTACCGGCAACTTGATAAAAAAATAAATTAGTATCGGCCTTACAGGAGATGGTAACAATTTTGTCAATTAATGA 3420

 AACTTTATTCGATTGCAATGGCCGTTGAACTATTTTTTTTATTTAATCATAGCCGGAATGTCCTCTACCATTGTTAAAACAGTAAATTACT
 Q F L S V N G A V Q Y E F L N T D A K C S I T V I K D N L S

 ATTTTGGATTACTGAATAGTGTTCATTATGATTCCCTCCTTTATTTAAGAATTAATCTTACATATAACATATAACATAATCAAGTTATTT 3510

 TAAAACCTAATGACTTATCACAAGGTAATACTAAGGAGGAATAAATTCCTAATTAGAATGTATATTGTATATTGTATTAGTTCAATAAA
 N Q I V S Y H E M ← HA34

 TTTGTAAACCTAAAATTTAAATATATCAAATTTTTATTAGTATGTTTACATAATTGATTATGGATATTTTCGTAAAAATGGCTTATTAAAA 3600

 AAACATTTGGATTTTAAATTTTATATAGTTTAAAAATAATCATACAAATGTATTAACATAACCTATAAAGCATTTTACCGAATAATTTT

 ← HA Promoter

ATTTAAAGGCAATTAGTTTATTTATAGTATAATAAAACAATAATATGTATATTATGGAGGGATAGTGGTAAATATGAATAAGTTGTTTTT 3690
 TAAATTTCCGTTAATCAAATAAATATCATATTATTTTGTATTATACATATAATACCTCCCTATCACCATTTATACTTATTCACAAAAA
 orf3 → M N K L F L
 ACAAATTAAAATGTTAAAAAATGACAACGAGGAGTTTCAAGAAATTTTAAAGCATTGTAAGAAAACTATAAAATATATTTACTAGAAAAATA 3780
 TGTTTAATTTTACAATTTTCTGTTGCTCCTCAAAGTTCTTTAAAAATTCGTAAACTTTTGTATTTATATAAATGATCTTTTAT
 Q I K M L K N D N E E F Q E I F K H F E K T I N I F T R K Y
 TAATATATATGATAATTACAATGATATTTTGTACCATTATGGTATACACTTAAAAAGTTGATTGAGCAATTTCAATACACAAAAATGA 3870
 ATTATATATACTATTAATGTTACTATAAAACATGGTAAATACCATATGTGAATTTTCAACTAACTCGTTAAAGTTATGTGTTTTACT
 N I Y D N Y N D I L Y H L W Y T L K K V D L S N F N T Q N D
 TTTAGAGAGATATATTAGTAGGACTTTAAAAAGATATTGCTTAGATATTTGCAATAAAAGAAAGATTGATAAGAAAATAATATATAATTC 3960
 AAATCTCTCTATATAATCATCCTGAAATTTTCTATAACGAATCTATAAACGTTATTTCTTTCTAACTATTCTTTTATTATATATTAAG
 L E R Y I S R T L K R Y C L D I C N K R K I D K K I I Y N S
 Hind III
 AGAAATTGCAGATAAGAAATTAAGCTTAATAGCAAATAGTTATTCAAGTTATTGAGAATTTGAATTTAATGATTTAATATCCATATTACC 4050
 TCTTTAACGCTATTCTTTAATTCGAATTATCGTTTATCAATAAGTTCAATAAGTCTTAACTTAAATTACTAAATTATAGGTATAATGG
 E I A D K K L S L I A N S Y S S Y S E F E F N D L I S I L P
 TGACGATCAAAAGAAAATTATATATATGAAATTTGTTGAAGATATTAAGGAGATAGATATAGCTAAAAAACTTAATATAAGTCGTCAATC 4140
 ACTGCTAGTTTTCTTTAATATATATACTTTAAACAACCTCTATAATTCCTCTATCTATATCGATTTTTTGAATTATATTGAGCAGTTAG
 D D Q K K I I Y M K F V E D I K E I D I A K K L N I S R Q S
 TGTATATAAAAAATAAAATAATGGCTTTAGAGAGATTAGAACCCATATTGAAAAATTAATTAATATGTAGTTTATATTTTAAAAAATTT 4230
 ACATATATTTTATTTTATTACCGAAATCTCTCTAATCTTGGGTATAACTTTTTTAATTAATTATACATCAAATATAAAAAATTTTTTAA
 V Y K N K I M A L E R L E P I L K K L I N M
 NTNH/BoNT/B^oPromoter
 TAGGTTTACAAAAAATAGTGTGGCTATGTTATATATAAATGATAAGAATATACTGAAAAATGTATCCAAATTTAAGGGGGCGTGTATAG 4320
 ATCCAAATGTTTTTATCACACCGATACAATATATATTTACTATTCTTATATGACTTTTTACATAGGTTTTAAATTCCTCCCGCACATATC
 TAAATAATTAAAAGTATGTGCGTTGAAATAAATTTAGGAGAGTGGTTAGATATGAATATAAATGACAACTTAAGTATAAATTCCTCCGGTA 4410
 ATTTATTAATTTTCATACACGCAACTTTATTTAAATCCTCTCACCAATCTATACTTATATTTACTGTTGAATTCATATTTAAGGGGCCAT
 NTNH → M N I N D N L S I N S P V
 GATAATAAAAAATGTTGTAGTAGTTAGAGCTAGAAAAACTGATACGGTTTTTAAGGCTTTTAAGGTTGCTCCCAATATTTGGGTGGCGCCA 4500
 CTATTATTTTACAACATCATCAATCTCGATCTTTTGTACTATGCCAAAAATTCGAAAAATTCACACGAGGGTTATAAACCCACCGCGGT
 D N K N V V V V R A R K T D T V F K A F K V A P N I W V A P

GAGAGATATTATGGCGAATCTTTGAGTATAGATGAAGAATATAAAGTTGATGGGGGAATATATGATTCTAATTTTCTTTCACAAGATAGT
CTCTCTATAATACCGCTTAGAACTCATATCTACTTCTTATATTTCAACTACCCCTTATATACTAAGATTAAAAGAAAGTGTCTATCA
E R Y Y G E S L S I D E E Y K V D G G I Y D S N F L S Q D S
Sca I
GAAAAAGATAAGTTCTTACAAGCCATTATTACTTTGTAAAAAGAATTAATAGTACTAACGCTGGGGAAAAGTTATTATCTTTGATTTCT
CTTTTTCTATTCAAGAATGTTGCGTAATAATGAAACAATTTTCTTAATTATCATGATTGCGACCCCTTTTCAATAATAGAACTAAAGA
E K D K F L Q A I I T L L K R I N S T N A G E K L L S L I S
ACAGCTATTCCATTTCTTATGGATATATAGGTGGAGGGTATTATGCACCTAATATGATTACTTTTGGATCAGCACCAAAATCTAATAAA
TGTCGATAAGGTAAAGGAATACCTATATATCCACCTCCCATAATACGTGGATTATACTAATGAAACCTAGTCGTGGTTTTAGATTATTT
T A I P F P Y G Y I G G G Y Y A P N M I T F G S A P K S N K
EcoR I Sca I
AAATTGAATTCCTTAATTTCAAGTACTATTCCATTTCTTATGCAGGATATAGAGAAACAAATTATCTTTCATCTGAAGATAATAAAAGT
TTTAACTTAAGAAATTAAGTTTCATGATAAGGTAAAGGAATACGTCCTATATCTCTTTGTTTAATAGAAAGTAGACTTCTATTATTTTCA
K L N S L I S S T I P F P Y A G Y R E T N Y L S S E D N K S
TTCTATGCATCTAATATAGTTATTTTTGGTCCAGGAGCAAACATAGTAGAAAACAATACTGTTTTTTATAAAAAGGAAGATGCAGAAAAT
AAGATACGTAGATTATATCAATAAAAACCGGTCCTCGTTTGTATCATCTTTGTTATGACAAAAATATTTTTCCTTCTACGTCTTTTA
F Y A S N I V I F G P G A N I V E N N T V F Y K K E D A E N
GGAATGGGAACAATGACTGAAATATGGTTCCAACCATTTCTAACCTATAAATATGACGAATTTTATATTGATCCTGCAATAGAATTAATA
CCTTACCCTTGTTACTGACTTTATACCAAGGTTGGTAAAGATTGGATATTTATACTGCTTAAATATAACTAGGACGTTATCTTAATTAT
G M G T M T E I W F Q P F L T Y K Y D E F Y I D P A I E L I
AAATGTTTAATAAAATCTCTTTATTTCTTATATGGTATAAAACCGAGTGATGATTTAGTTATTCCATATAGATTAAGAAGTGAATTAGAG
TTTACAAATTATTTTAGAGAAATAAAGAATATACCATATTTGGCTCACTACTAAATCAATAAGGTATATCTAATTCTTCACTTAATCTC
K C L I K S L Y F L Y G I K P S D D L V I P Y R L R S E L E
AATATAGAATACTCACAATTGAATATAGTTGATTACTAGTATCTGGAGGCATTGATCCTAAATTTATAAATACAGATCCATATTGGTTT
TTATATCTTATGAGTGTTAACTTATATCAACTAAATGATCATAGACCTCCGTAAGTAAATATTTATGTCTAGGTATAACCAAA
N I E Y S Q L N I V D L L V S G G I D P K F I N T D P Y W F
ATAGATAATTATTTCTCAAATGCAAAAAAAGTGTTTGAAGATCATAGGAATATTTATGAAACAGAAATTGAAGGAAATAATGCCATTGGT
TATCTATTAATAAAGAGTTTACGTTTTTTTACAAACTTCTAGTATCCTTATAAATACTTTGTCTTTAACTTCCTTTATTACGGTAACCA
I D N Y F S N A K K V F E D H R N I Y E T E I E G N N A I G
Taq I
AATGATATAAAATTGAGATTAAAACAAAAGTTTCGAATCAATATCAATGATATATGGGAATTAAATTTAAATTATTTCTCTAAAGAGTTT
TTACTATATTTTAACTCTAATTTTGTTCGAAAGCTTAGTTATAGTTACTATATACCCTTAATTTAAATTTAATAAAGAGATTTCTCAA
N D I K L R L K Q K F R I N I N D I W E L N L N Y F S K E F

AGCATTATGATGCCAGATAGATTTAATAATGCACTTAAACATTTTTATAGAAAACAATACTACAAAATAGATTATCCAGAAAATTATAGT
 ----- 5490
 TCGTAATACTACGGTCTATCTAAATTATTACGTGAATTTGTAAAAATATCTTTTGTATGATGTTTTATCTAATAGGTCTTTTAATATCA
 S I M M P D R F N N A L K H F Y R K Q Y Y K I D Y P E N Y S

ATAAATGGTTTTGTTAATGGTCAAATTAATGCTCAATTATCTTTATCAGATAGAAATCAAGATATTATAAATAAACCTGAAGAAATAATT
 ----- 5580
 TATTTACCAAACAATTACCAGTTTAATTACGAGTTAATAGAAATAGTCTATCTTTAGTTCTATAATATTTATTTGGACTTCTTTATTAA
 I N G F V N G Q I N A Q L S L S D R N Q D I I N K P E E I I

AATTTATTAAATGGAAATAATGTTTCATTAATGAGAAGTAATATTTATGGTGATGGATTAAAAAGCACTGTAGATGATTTTTACAGTAAT
 ----- 5670
 TTAAATAATTTACCTTTATTACAAAGTAATTACTCTTCATTATAAATACCACTACCTAATTTTTTCGTGACATCTACTAAAAATGTCATTA
 N L L N G N N V S L M R S N I Y G D G L K S T V D D F Y S N

TATAAAATCCCATATAATAGAGCCTATGAATATCATTTTAATAATTCAAATGATTCTTCTTTAGATAATGTTAACATTGGAGTAATAGAC
 ----- 5760
 ATATTTTAGGGTATATTATCTCGGATACTTATAGTAAAATTATTAAGTTTACTAAGAAGAAATCTATTACAATTGTAACCTCATTATCTG
 Y K I P Y N R A Y E Y H F N N S N D S S L D N V N I G V I D

AATATTCCAGAGATTATAGATGTAAATCCTTATAAGGAAAATTGTGATAAGTTTTCCCGGTACAGAAAATTACAAGTACTAGAGAAATT
 ----- 5850
 TTATAAGGTCTCTAATATCTACATTTAGGAATATTCCTTTTAACACTATTCAAAGTGGCCATGTCTTTTAATGTTTCATGATCTCTTTAA
 N I P E I I D V N P Y K E N C D K F S P V Q K I T S T R E I

Hae III
 AATACAAATATACCATGGCCTATAAATTATTTACAAGCTCAAATACTAACAATGAAAAATTTAGTTTATCCTCAGATTTTGTAGAAGTA
 ----- 5940
 TTATGTTTATATGGTACCGGATATTTAATAAATGTTTCGAGTTTTATGATTGTTACTTTTTAAATCAAATAGGAGTCTAAAACATCTTCAT
 N T N I P W P I N Y L Q A Q N T N N E K F S L S S D F V E V

GTTTCTTCTAAAGATAAATCTTTAGTGTATTCTTTCTTATCTAATGTAATGTTTTATTTAGATTCCATAAAGGATAATAGTCCTATTGAT
 ----- 6030
 CAAAGAAGATTTCTATTTAGAAATCACATAAGAAAGAATAGATTACATTACAAAATAAATCTAAGGTATTTCTTATTATCAGGATAACTA
 V S S K D K S L V Y S F L S N V M F Y L D S I K D N S P I D

ACAGATAAAAAATATTATTTATGGTTAAGAGAGATTTTTAGAAATTATTCTTTTGATATTACTGCAACTCAAGAAATTAATACTGATTGC
 ----- 6120
 TGTCTATTTTTTATAATAAATACCAATTCTCTCTAAAAATCTTTAATAAGAAAACTATAATGACGTTGAGTTCTTTAATTATGACTAAGC
 T D K K Y Y L W L R E I F R N Y S F D I T A T Q E I N T D C

GGTATTAATAAAGTAGTAACTTGGTTTGGAAAAGCATTAAATATTTTAAATACATCCGATTCTTTTGTAGAAGAATTTCAAATTTAGGA
 ----- 6210
 CCATAATTATTTTCATCATTGAACCAAACCTTTTCGTAATTTATAAAATTTATGTAGGCTAAGAAAACATCTTCTTAAAGTTTTAAATCCT
 G I N K V V T W F G K A L N I L N T S D S F V E E F Q N L G

Cla I
 CCAATTTCACTTATTAATAAAAAAGAAAATTTAAGTATGCCTAAAATAGAAATCGATGAAATACCTAATAGTATGTTGAATCTTTCATTT
 ----- 6300
 GGTAAAGTGAATAATTATTTTTCTTTTAAATTCATACGGATTTTATCTTTAGCTACTTTATGGATTATCATACAACCTTAGAAAGTAAA
 P I S L I N K K E N L S M P K I E I D E I P N S M L N L S F

AAAGATTTAAGTGAAAATTTATTTAATATATTCTCTAAAAACAACTCTTATTTTGAAAAATCTACTATGATTTCTTAGATCAATGGTGG 6390

TTTCTAAATTCACCTTTTAAATAAATTATATAAGAGATTTTGTGAGAATAAACTTTTTTAGATGATACTAAAGAATCTAGTTACCACC
K D L S E N L F N I F S K N N S Y F E K I Y Y D F L D Q W W

ACACAATATTATAGTCAATATTTTGATTTAATTTGTATGGCTAAAAGATCAGTGTTAGCTCAAGAAAGTTAATAAAAAAATAATACAA 6480

TGTGTTATAATATCAGTTATAAACTAAATTAAACATACCGATTTTCTAGTCACAATCGAGTTCTTCAAATTATTTTTTTTATTATGTT
T Q Y Y S Q Y F D L I C M A K R S V L A Q E S L I K K I I Q

AAAAAATTGAGTTATTTAATAGGAAATCTAATATATCATCTGATAACTTAGCATTGATGAATCTGACAACAACAAATACATTAAGAGAT 6570

TTTTTTAACTCAATAAATTATCCTTTAAGATTATATAGTAGACTATTGAATCGTAACTACTTAGACTGTTGTTGTTTATGTAATTCTCTA
K K L S Y L I G N S N I S S D N L A L M N L T T T N T L R D

ATTTCAAACGAATCACAAATAGCAATGAACAATGTAAACAATTTTTTAAACAATGTTGCTATATGTGTTTTTCAAATAATATATATCCT 6660

TAAAGTTTGCTTAGTGTTTATCGTTACTTGTTACATTTGTTAAAAAATTTGTTACAACGATATACACAAAAGTTTGATTATATATAGGA
I S N E S Q I A M N N V N N F L N N V A I C V F Q T N I Y P

AAGTTTATTTCTTTTATGGAACAATGTATTAATAATATAAATAAGAATACAAGAGAATTTATACAAAAATGTACCAATATTACTGAAAAT 6750

TTCAAATAAAGAAAATACCTTGTTACATAATTATTATATTTATTTCTTATGTTCTCTTAAATATGTTTTTACATGGTTATAATGACTTTTA
K F I S F M E Q C I N N I N K N T R E F I Q K C T N I T E N

GAAAAATTACAATTGATTAACCAAAATATTTTTAGTAGCTTAGATTTTGATTTTTTAAACATTGAAAATTTGAAAAGCTTATTTAATTCA 6840

CTTTTTAATGTTAACTAATTGGTTTTTATAAAAATCATCGAATCTAAACTAAAAAATTTGTAACCTTTTAACTTTTCGAATAAATTAAGT
E K L Q L I N Q N I F S S L D F D F L N I E N L K S L F N S

GAGACAGGATTACTTATAAAGGAGGAACTTCACCCTATGAACTAGTGTTATATGCTTTTCAGGAACCAGGTAATAATGCTATCGGAGAT 6930

CTCTGTCCTAATGAATATTTCTCCTTTGAAGTGGGATACTTGATCACAATATACGAAAAGTCCTTGGTCCATTATTACGATAGCCTCTA
E T G L L I K E E T S P Y E L V L Y A F Q E P G N N A I G D

GCATCTGGTAAAAATACATCAATAGAATATTCTAAGGACATAGGTTTAGTTTATGGAATAAATAGTGATGCATTATATTTAAATGGATCT 7020

CGTAGACCATTTTTATGATGTTATCTTATAAGATTCCTGTATCCAAATCAAATACCTTATTTATCACTACGTAATATAAATTTACCTAGA
A S G K N T S I E Y S K D I G L V Y G I N S D A L Y L N G S

AATCAAAGTATAAGTTTTTCTAATGATTTCTTTGAAAATGGATTAACCTAATAGTTTTTCAATTTATTTTTGGTTGAGAAATTTGGGCAAA 7110

TTAGTTTCATATTCAAAAAGATTACTAAAGAACTTTTACCTAATTGATTATCAAAAAGTTAAATAAAAACCAACTCTTTAAACCCGTTT
N Q S I S F S N D F F E N G L T N S F S I Y F W L R N L G K

GATACTATTAAATCTAAGCTAATAGGTAGTAAGGAAGATAATTGTGGTTGGGAAATTTATTTTCAAGATACTGGGTTGGTTTTTAATATG 7200

CTATGATAATTTAGATTGATTATCCATCATTCTTCTATTAACACCAACCCTTTAAATAAAAGTTCTATGACCCAACCAAAAATTATAC
D T I K S K L I G S K E D N C G W E I Y F Q D T G L V F N M

Hind III

APPENDIX IV

COMPARATIVE ALIGNMENT OF THE NTNII GENE FROM THE TYPE B NEUROTOXIN LOCUS IN C.

ATAGATTCTAATGGAAACGAGAAGAATATATATCTATCTGATGTTTCTAATAATAGTTGGCACTATATACTATATCTGTAGATCGTTTA	7290
TATCTAAGATTACCTTTGCTCTTCTTATATATAGATAGACTACAAAGATTATTATCAACCGTGATATATTGATATAGACATCTAGCAAAT	
I D S N G N E K N I Y L S D V S N N S W H Y I T I S V D R L	
AAAGAACAATTATTAATATTTATTGATGATAATTTAGTGGCTAATGAAAGTATTAAAGAAATTTTAAATATCTATTCAAGTAATACAATT	7380
TTTCTTGTTAATAATTATAAATAACTACTATTAAATCACCGATTACTTTTCATAATTTCTTTAAATTTATAGATAAGTTCATTATGTTAA	
K E Q L L I F I D D N L V A N E S I K E I L N I Y S S N T I	
TCTTTAGTTAATGAAAATAATCCAATATATGTTGAAGGATTATCTATTTTGAATAGATCTATTACAAGTGAAGAAGTTGTAAATAATTAT	7470
AGAAATCAATTACTTTTATTAGGTTATATACAACCTTCCTAATAGATAAACTTATCTAGATAATGTTCACTTCTTCAACATTTATTAATA	
S L V N E N N P I Y V E G L S I L N R S I T S E E V V N N Y	
TTTACTTATTTAAATAATTCATATATAAGGGATATTTCTGGAGAAAGATTAGAATACAATAAAACCTATGAACTATACAATTATGTATTC	7560
AAATGAATAAATTTATTAAGTATATATTCCTATAAAGACCTCTTTCTAATCTTATGTTATTTTGGATACTTGATATGTTAATACATAAG	
F T Y L N N S Y I R D I S G E R L E Y N K T Y E L Y N Y V F	
CCAGAAAGCTCGTTATATGAAGTTACTGAAAACAATAATATATATTTGTCAATTAAAAATACAAATAACCTTAATATTCAAGGTGCTAAG	7650
GGTCTTTTCGAGCAATATACTTCAATGACTTTTGTATTATATATAAACAGTTAATTTTTATGTTTATTGGAATTATAAGTTCCACGATTC	
P E S S L Y E V T E N N N I Y L S I K N T N N L N I Q G A K	
TTTAAATTAATAAATATTGACGCAAACAAGCAATATGTTCAAAAGTGGGATGAAGGAGTAGTTTGTATTATTAGGAGATGAAGAAAAATAT	7740
AAATTTAATTATTTATAACTGCGTTTGTTCGTTATACAAGTTTTCACCTACTTCTCATCAAACAAATAATCCTCTACTTCTTTTTATA	
F K L I N I D A N K Q Y V Q K W D E G V V C L L G D E E K Y	
GTAGATATATCCTCTGAAAATAATAGAATCCAATTAGTTAGCTCTAAAGATACTGCAAAGAGAATAATATTTAATAACGATATATTTAGG	7830
CATCTATATAGGAGACTTTTATTATCTTAGGTTAATCAATCGAGATTTCTATGACGTTTCTCTTATTATAAATTATTGCTATATAAATCC	
V D I S S E N N R I Q L V S S K D T A K R I I F N N D I F R	
CCTAATTGCCTTACATTTGCTTATAATAATAAATATTTATCTTTATCCTTAAGAGATAGAAATTATAATTGGATGATATGTAATAATAAT	7920
GGATTAACGGAATGTAAACGAATATTATTATTTATAAATAGAAATAGGAATTCTCTATCTTTAATATTAACCTACTATACATTATTATTA	
P N C L T F A Y N N K Y L S L S L R D R N Y N W M I C N N N	
GACAATATACCTAAGGCTGCACATTTATGGGCATTAAAAGGGATATAAACTTAAAATAAGGAGGAGAATATTTATGCCAGTTACAATAAA	8010
CTGTTATATGGATTCCGACGTGTAAATACCCGTAATTTTCCCTATATTTGAATTTTATTCCTCCTCTTATAAATACGGTCAATGTTATTT	
D N I P K A A H L W A L K G I .	
BoNT/B° → M P V T I N	

APPENDIX III. Annotated nucleotide sequence of neurotoxin complex operons 5' to the silent type B neurotoxin in *C. botulinum* NCTC 2916

Open reading frames are translated below the corresponding sequence, Red font indicates translation from the sense strand whilst blue indicates translation from the reverse strand. Putative ribosome binding sites are shown in red, selected restriction sites; = and Promoter sequences of the Haemagglutinin operon and NTNII/BoNT/B°; ==

APPENDIX IV

COMPARATIVE ALIGNMENT OF THE NTN^H GENE FROM THE TYPE B NEUROTOXIN LOCUS IN C. BOTULINUM TYPE A, STRAIN NCTC 2916, WITH NTN^H GENES FROM OTHER SEROTYPES

	A1	A2	A3	SB	B	C	E	F
241	G	C	C	A	A	T	T	A
241	G	C	C	A	A	T	T	A
241	G	C	C	A	A	T	T	A
241	G	C	C	A	A	T	T	A
241	G	C	C	A	A	T	T	A
241	G	C	C	A	A	T	T	A
238	G	C	C	A	A	T	T	A
241	G	C	C	A	A	T	T	A

	A1	A2	A3	SB	B	C	E	F
301	T	T	G	A	T	T	G	A
301	T	T	G	A	T	T	G	A
301	T	T	G	A	T	T	G	A
301	T	T	G	A	T	T	G	A
301	T	T	G	A	T	T	G	A
301	T	T	G	A	T	T	G	A
298	T	T	G	A	T	T	G	A
301	T	T	G	A	T	T	G	A

	A1	A2	A3	SB	B	C	E	F
361	A	A	T	A	T	G	A	T
361	A	A	T	A	T	G	A	T
345	-	-	-	-	-	-	-	-
361	A	A	T	A	T	G	A	T
361	A	A	T	A	T	G	A	T
361	A	A	T	A	T	G	A	T
342	-	-	-	-	-	-	-	-
345	-	-	-	-	-	-	-	-

	A1	A2	A3	SB	B	C	E	F
421	A	G	T	A	C	T	A	T
421	A	G	T	A	C	T	A	T
346	-	-	-	-	-	-	-	-
421	A	G	T	A	C	T	A	T
421	A	G	T	A	C	T	A	T
421	A	G	T	A	C	T	A	T
343	-	-	-	-	-	-	-	-
346	-	-	-	-	-	-	-	-

	A1	A2	A3	SB	B	C	E	F
481	A	A	A	A	A	A	A	A
481	A	A	A	A	A	A	A	A
382	A	A	A	A	A	A	A	A
481	A	A	A	A	A	A	A	A
481	A	A	A	A	A	A	A	A
481	A	A	A	A	A	A	A	A
379	A	A	A	A	A	A	A	A
382	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
541	A	A	A	A	A	A	A	A
541	A	A	A	A	A	A	A	A
442	A	A	A	A	A	A	A	A
541	A	A	A	A	A	A	A	A
541	A	A	A	A	A	A	A	A
541	A	A	A	A	A	A	A	A
439	A	A	A	A	A	A	A	A
442	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
601	A	A	A	A	A	A	A	A
601	A	A	A	A	A	A	A	A
502	A	A	A	A	A	A	A	A
601	A	A	A	A	A	A	A	A
601	A	A	A	A	A	A	A	A
601	A	A	A	A	A	A	A	A
499	A	A	A	A	A	A	A	A
502	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
661	G	A	A	A	A	A	A	A
661	G	A	A	A	A	A	A	A
562	G	A	A	A	A	A	A	A
661	G	A	A	A	A	A	A	A
661	G	A	A	A	A	A	A	A
661	G	A	A	A	A	A	A	A
559	G	A	A	A	A	A	A	A
562	G	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
961	A	A	G	A	A	A	A	G
961	A	A	A	A	A	A	A	A
862	A	A	A	A	A	A	A	A
961	A	A	A	A	A	A	A	A
961	A	A	A	A	A	A	A	A
961	A	A	A	A	A	A	A	A
859	A	A	A	A	A	A	A	A
862	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
1021	A	A	A	A	A	A	A	A
1021	A	A	A	A	A	A	A	A
922	A	A	A	A	A	A	A	A
1021	A	A	A	A	A	A	A	A
1021	A	A	A	A	A	A	A	A
1021	A	A	A	A	A	A	A	A
919	A	A	A	A	A	A	A	A
922	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
1081	A	A	A	A	A	A	A	A
1081	A	A	A	A	A	A	A	A
982	A	A	A	A	A	A	A	A
1081	A	A	A	A	A	A	A	A
1081	A	A	A	A	A	A	A	A
1081	A	A	A	A	A	A	A	A
979	A	A	A	A	A	A	A	A
982	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
1141	C	A	A	A	A	A	A	A
1141	C	A	A	A	A	A	A	A
1042	C	A	A	A	A	A	A	A
1141	C	A	A	A	A	A	A	A
1141	C	A	A	A	A	A	A	A
1141	C	A	A	A	A	A	A	A
1039	C	A	A	A	A	A	A	A
1042	C	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
1201	G	A	A	T	T	T	T	T
1201	G	A	A	T	T	T	T	T
1102	T	T	T	T	T	T	T	T
1201	G	A	A	T	T	T	T	T
1201	G	A	A	T	T	T	T	T
1201	A	A	A	T	T	T	T	T
1099	T	T	T	T	T	T	T	T
1102	T	T	T	T	T	T	T	T

	A1	A2	A3	SB	B	C	E	F
1258	G	G	G	G	G	G	G	G
1258	G	G	G	G	G	G	G	G
1159	G	G	G	G	G	G	G	G
1258	G	G	G	G	G	G	G	G
1258	G	G	G	G	G	G	G	G
1258	G	G	G	G	G	G	G	G
1156	G	G	G	G	G	G	G	G
1162	G	G	G	G	G	G	G	G

	A1	A2	A3	SB	B	C	E	F
1318	A	G	A	G	C	C	C	C
1318	A	G	A	G	C	C	C	C
1219	A	G	A	G	C	C	C	C
1318	A	G	A	G	C	C	C	C
1318	A	G	A	G	C	C	C	C
1318	A	G	A	G	C	C	C	C
1216	C	T	A	A	T	T	T	T
1222	A	T	A	A	T	T	T	T

	A1	A2	A3	SB	B	C	E	F
1378	G	G	A	A	T	T	T	T
1378	G	G	A	A	T	T	T	T
1276	A	A	A	A	T	T	T	T
1378	G	G	A	A	T	T	T	T
1378	G	G	A	A	T	T	T	T
1378	G	G	A	A	T	T	T	T
1267	G	G	A	A	T	T	T	T
1273	G	G	A	A	T	T	T	T

	A1	A2	A3	SB	B	C	E	F
1432	T	G	T	G	A	T	A	T
1432	T	G	T	G	A	T	A	T
1336	T	G	T	G	A	T	A	T
1432	T	G	T	G	A	T	A	T
1432	T	G	T	G	A	T	A	T
1432	T	G	T	G	A	T	A	T
1327	T	G	T	G	A	T	A	T
1333	T	G	T	G	A	T	A	T

	A1	A2	A3	SB	B	C	E	F
1492	C	C	C	C	C	C	C	C
1492	C	C	C	C	C	C	C	C
1396	C	C	C	C	C	C	C	C
1492	C	C	C	C	C	C	C	C
1492	C	C	C	C	C	C	C	C
1492	C	C	C	C	C	C	C	C
1387	C	C	C	C	C	C	C	C
1393	C	C	C	C	C	C	C	C

	A1	A2	A3	SB	B	C	E	F
1552	T	C	C	C	C	C	C	C
1552	T	C	C	C	C	C	C	C
1456	T	C	C	C	C	C	C	C
1552	T	C	C	C	C	C	C	C
1552	T	C	C	C	C	C	C	C
1552	T	C	C	C	C	C	C	C
1447	T	C	C	C	C	C	C	C
1453	T	C	C	C	C	C	C	C

	A1	A2	A3	SB	B	C	E	F
1612	A	A	A	A	A	A	A	A
1612	A	A	A	A	A	A	A	A
1516	A	A	A	A	A	A	A	A
1612	A	A	A	A	A	A	A	A
1612	A	A	A	A	A	A	A	A
1612	A	A	A	A	A	A	A	A
1504	A	A	A	A	A	A	A	A
1513	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
2151	T	G	G	A	T	G	A	A
2151	T	G	G	A	T	G	A	A
2055	T	G	G	A	T	G	A	A
2151	T	G	G	A	T	G	A	A
2148	T	G	G	A	T	G	A	A
2148	T	G	G	A	T	G	A	A
2043	T	G	G	A	T	G	A	A
2052	T	G	G	A	T	G	A	A

	A1	A2	A3	SB	B	C	E	F
2211	A	T	T	T	T	T	T	T
2211	A	T	T	T	T	T	T	T
2115	A	T	T	T	T	T	T	T
2211	A	T	T	T	T	T	T	T
2208	A	T	T	T	T	T	T	T
2208	A	T	T	T	T	T	T	T
2103	A	T	T	T	T	T	T	T
2112	A	T	T	T	T	T	T	T

	A1	A2	A3	SB	B	C	E	F
2271	T	C	A	A	A	A	A	A
2271	T	C	A	A	A	A	A	A
2175	T	C	A	A	A	A	A	A
2271	T	C	A	A	A	A	A	A
2268	T	C	A	A	A	A	A	A
2268	T	C	A	A	A	A	A	A
2163	T	C	A	A	A	A	A	A
2172	T	C	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
2331	T	A	A	A	A	A	A	A
2331	T	A	A	A	A	A	A	A
2235	T	A	A	A	A	A	A	A
2331	T	A	A	A	A	A	A	A
2328	T	A	A	A	A	A	A	A
2328	T	A	A	A	A	A	A	A
2223	T	A	A	A	A	A	A	A
2232	T	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
2631	A	T	T	T	A	T	A	A
2631	A	T	T	T	A	T	A	A
2535	A	T	T	T	A	T	A	A
2631	A	T	T	T	A	T	A	A
2628	A	T	T	T	A	T	A	A
2628	A	T	T	T	A	T	A	A
2523	A	T	T	T	A	T	A	A
2532	A	T	T	T	A	T	A	A

	A1	A2	A3	SB	B	C	E	F
2691	A	T	T	T	A	T	A	A
2691	A	T	T	T	A	T	A	A
2595	A	T	T	T	A	T	A	A
2691	A	T	T	T	A	T	A	A
2688	A	T	T	T	A	T	A	A
2688	A	T	T	T	A	T	A	A
2583	A	T	T	T	A	T	A	A
2592	A	T	T	T	A	T	A	A

	A1	A2	A3	SB	B	C	E	F
2751	A	T	T	T	A	T	A	A
2751	A	T	T	T	A	T	A	A
2655	A	T	T	T	A	T	A	A
2751	A	T	T	T	A	T	A	A
2748	A	T	T	T	A	T	A	A
2748	A	T	T	T	A	T	A	A
2643	A	T	T	T	A	T	A	A
2652	A	T	T	T	A	T	A	A

	A1	A2	A3	SB	B	C	E	F
2811	A	T	T	T	A	T	A	A
2811	A	T	T	T	A	T	A	A
2715	A	T	T	T	A	T	A	A
2811	A	T	T	T	A	T	A	A
2808	A	T	T	T	A	T	A	A
2808	A	T	T	T	A	T	A	A
2703	A	T	T	T	A	T	A	A
2712	A	T	T	T	A	T	A	A

	A1	A2	A3	SB	B	C	E	F
3112	A	A	A	A	A	A	A	A
3112	A	A	A	A	A	A	A	A
3016	C	C	C	G	G	C	G	C
3112	T	T	T	T	T	T	T	T
3109	A	A	A	A	A	A	A	A
3109	T	T	T	T	T	T	T	T
3004	C	C	C	C	C	C	C	C
3013	A	A	A	A	A	A	A	A

	A1	A2	A3	SB	B	C	E	F
3172	T	T	T	T	T	T	T	T
3172	T	T	T	T	T	T	T	T
3076	C	C	C	C	C	C	C	C
3172	T	T	T	T	T	T	T	T
3169	T	T	T	T	T	T	T	T
3169	T	T	T	T	T	T	T	T
3064	T	T	T	T	T	T	T	T
3073	T	T	T	T	T	T	T	T

	A1	A2	A3	SB	B	C	E	F
3232	T	T	T	T	T	T	T	T
3232	T	T	T	T	T	T	T	T
3136	T	T	T	T	T	T	T	T
3232	T	T	T	T	T	T	T	T
3229	T	T	T	T	T	T	T	T
3229	T	T	T	T	T	T	T	T
3124	T	T	T	T	T	T	T	T
3133	T	T	T	T	T	T	T	T

	A1	A2	A3	SB	B	C	E	F
3292	A	A	A	A	A	A	A	A
3292	A	A	A	A	A	A	A	A
3196	T	T	T	T	T	T	T	T
3292	A	A	A	A	A	A	A	A
3289	T	T	T	T	T	T	T	T
3289	T	T	T	T	T	T	T	T
3184	A	A	A	A	A	A	A	A
3193	A	A	A	A	A	A	A	A

3579.		A1
3579.		A2
3483.		A3
3592	A T A.	SB
3589	A T A	B
3588	T.	C
3484	A T A.	E
3493	G T A.	F

APPENDIX IV:
COMPARATIVE ALIGNMENT OF THE NTN_H GENE FROM THE TYPE B NEUROTOXIN LOCUS IN C.
BOTULINUM TYPE A, STRAIN NCTC2916, WITH NTN_H GENES FROM OTHER SEROTYPES

38

- A1: Accession No. L42537, "chimeric" NTN_H gene from the type A locus, strain NCTC 2916
- A2: Accession No. X92973, NTN_H gene from "classical" type A₁ strain ATCC 3502
- A3: Accession No.X87848, NTN_H gene from type A₂ strain Kyoto-F
- SB: This study, NTN_H gene from silent type B neurotoxin locus in type A strain NCTC 2916
- B: Accession No. X78230, NTN_Hgene from proteolytic type B strain
- C: Accession No. X72793, NTN_H gene from type C Stockholm strain (Very similar to type D)
- E: Accession No. P46082, NTN_H gene from type E Mashike strain
- F: Accession No. X71086, NTN_H gene from type F Langeland strain

APPENDIX V

OLIGONUCLEOTIDE PRIMERS

M13 Forward (-20) Primer

5'-GTAAAACGACGGCCAG-3'

M13 Reverse Primer

5'-CAGGAAACAGCTATGAC-3'

B2676

5'-GTAAATAATTATTTTCTTATT-3'

B2677

5'-CTCTCGCAAATGGAGGCTCCATC-3'

B2708

5'-TTTTTTAGATCAATGGTGGAC-3'

B2709

5'-ATTCTCCGGTGGATAATAAGAATG-3'

3237

5'-TCGAAATATTATAGAGGGG-3'